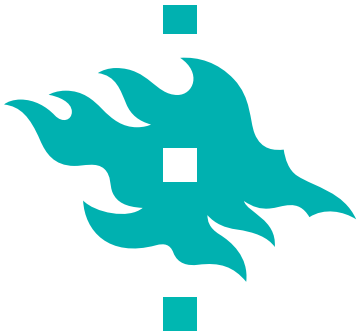

DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

MAIJU RINNE

MOLECULAR EVOLUTION OF G PROTEIN-COUPLED RECEPTORS — INSIGHTS INTO THE OREXIN SYSTEM

DIVISION OF PHARMACEUTICAL CHEMISTRY AND TECHNOLOGY
FACULTY OF PHARMACY
DOCTORAL PROGRAMME IN INTEGRATIVE LIFE SCIENCE
UNIVERSITY OF HELSINKI

Division of Pharmaceutical Chemistry and Technology
Faculty of Pharmacy
University of Helsinki
Finland

Molecular evolution of G protein-coupled receptors – insights into the orexin system

Maiju K. Rinne

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Pharmacy, University of Helsinki, for public examination in the Auditorium 1041, Biocentre 2, on September 27th 2019, at 12 o'clock.

Helsinki 2019

Supervisors Docent Henri Xhaard, PhD
Division of Pharmaceutical Chemistry and Technology
Faculty of Pharmacy
University of Helsinki, Finland

Professor Jyrki Kukkonen, PhD
Department of Veterinary Biosciences
Faculty of Veterinary Medicine
University of Helsinki, Finland

Reviewers Docent Juha Savinainen, PhD
Institute of Biomedicine
University of Eastern Finland, Finland

Docent Tiina Salminen, PhD
Structural Bioinformatics Laboratory
Biochemistry, Faculty of Science and Engineering
Åbo Akademi University, Finland

Opponent Senior Scientist Xavier Deupi i Corral, PhD
Laboratory for Scientific Computing and Modelling
Paul Scherrer Institute, Switzerland

ISBN 978-951-51-5468-2 (paperback)
ISBN 978-951-51-5469-9 (PDF, http://ethesis.helsinki.fi)

Published in the DSHealth series *Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis*
ISSN 2342-3161 (print)
ISSN 2342-317X (online)

Hansaprint, Vantaa, Finland, 2019

Abstract

The diversity of G protein-coupled receptors is a fundamental element in this thesis. GPCRs are the largest family of membrane proteins, with about 800 in humans. While the human GPCR repertoire is well described, in other species, especially in non-mammals, GPCRs are not tracked to the individual subtype level in large-scale genomic studies. The diversity of GPCRs in non-human vertebrates was studied in the first publication. The study classified 142 rhodopsin-like non-olfactory GPCRs without human orthologue, 69 of which were reported for the first time. The study also points out inconsistencies in the GPCR nomenclature system and reveals a pool of yet-to-be studied receptors. Understanding the repertoire of GPCRs in non-human species might also be useful for many areas of science, such as pharmacology, ecotoxicology and evolutionary biology.

For more insight into the orexin system, two small-molecule orexin receptor agonists were pharmacologically characterised in the second and third publications. Nag26 was confirmed to be a potent and almost full agonist of orexin receptors, but the selectivity for orexin receptor type-2 was not as strong as reported (20-fold against 70-fold). Yan7874 was also confirmed to be an orexin receptor agonist, but only partial and weak with high off-target activity. Neither of these compounds is likely to be suitable for further drug development as such. These studies also display the challenge in the development of small-molecule agonists for peptide-bound GPCRs.

Finally, *Ciona intestinalis* putative orexin receptor was studied, and its functionality was verified in recombinant cells. Homology models of *C. intestinalis* orexin receptor revealed a highly similar binding cavity as in the human OX_2 orexin receptor. Thus, the functionality of the receptor was studied in Ca^{2+} elevation assay, and the receptor was verified to be functional and binding to human orexin peptides. Further database mining resulted in the identification of putative *C. intestinalis* prepro-orexin and orexin peptide that was shown to bind to a plasma membrane of *C. intestinalis* orexin receptor-expressing cells. Solving the function of the orexin system in distinct species such as *C. intestinalis* might be interesting from an evolutionary point of view but also essential in extending the knowledge of the orexin system in humans.

This thesis was conducted collaboratively with the Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy and Department of Veterinary Biosciences, Faculty of Veterinary Medicine, at the University of Helsinki.

Tiivistelmä

G-proteiinikytkentäisten reseptorien moninaisuus on keskiössä tässä väitöskirjatyössä. G-proteiinikytkentäiset reseptorit ovat suurin yhtenäinen kalvoreseptorien ryhmä, ja ihmisellä on noin 800 G-proteiinikytkentäistä reseptoria. Toisin kuin ihmisen reseptorien kohdalla, muiden lajien edustajien (varsinkin nisäkkäiden ulkopuolelta) reseptorijoukkoa on harvoin jäljitetty alatyyppitasolle asti laajoissa genomitutkimuksissa.

Tämän väitöskirjatyön ensimmäinen julkaisu perehtyy G-proteiinikytkentäisten reseptorien moninaisuuteen selkärankaisilla lajeilla. Tutkimus luokittelee 142 rodopsiini-reseptorien ryhmään kuuluvaa G-proteiinikytkentäistä reseptoria ilman humaania ortologia. Näistä reseptoreista 69 esitetään tässä työssä ensimmäistä kertaa. Lisäksi tutkimus osoittaa reseptorien epäjohdonmukaisuuksia nimeämissysteemissä, sekä esittelee vielä tutkimattomia reseptoreja. Laajempi ymmärrys eri lajien reseptorirepertuaarista voi olla hyödyllinen monilla tieteenaloilla, kuten farmakologiassa, ekotoksikologiassa ja evoluutiobiologissa.

Tässä väitöskirjatyössä tutkittiin myös ihmisen G-proteiinikytkentäisiä reseptoreja. Kahden pienmolekyylioreksiinireseptoriagonistin ominaisuudet luokiteltiin farmakologisesti tämän väitöskirjan toisessa ja kolmannessa julkaisussa. Tutkimus vahvisti pienmolekyyli Nag26:n olevan potentti, lähes täysagonisti ja selektiivinen tyypin-2 oreksiinireseptorille, joskaan ei yhtä voimakkaasti selektiivinen kuin aikaisemmin on raportoitu. Pienmolekyyli Yan7874:n vahvistettiin olevan myös oreksiinireseptoriagonisti, mutta vain osittainen ja heikko. Lisäksi, Yan7874 indusoi soluissa epäspesifisiä vaikutuksia. On siis epätodennäköistä, että kumpikaan näistä yhdisteistä tarjoaisi sellaisenaan potentiaalia pidemmälle lääkekehitykselle. Nämä tutkimukset osoittavat osaltaan pienmolekyyliagonistien kehityksen haasteita peptideihin sitoutuviin G-proteiinikytkentäistiin reseptoreihin.

Lopuksi, tässä työssä tutkittiin selkärangattoman *Ciona intestinalis* –lajin putatiivista oreksiinireseptoria, ja sen toiminnallisuus vahvistettiin ilmentämällä sitä rekombinanteissa soluissa. Homologiamallien perusteella reseptorin sitoutumistaskun osoitettiin muistuttavan ihmisen tyyppi-2 oreksiinireseptorin sitoutumistaskua. Oreksiinireseptorin toimivuus vahvistettiin reseptori-riippuvaisella solunsisäisen kalsiumin mobilisaatiolla, ja sitovan humaaneja oreksiinipeptidejä. Lisäksi tietokantalouhinta johti putatiivisen *C. intestinalis* oreksiini-prekursorin ja oreksiinipeptidin tunnistamiseen. Tässä väitöskirjatyössä tunnistettiin ensimmäistä

kertaa *C. intestinalis* –lajin oreksiini-prekursori ja osoitettiin, että putatiivinen oreksiinipeptidi sitoutuu *C. intestinalis* oreksiinireseptoria expressoivien solujen solukalvolle. Oreksiinisysteemin toiminnallisuus evolutiivisesti etäisillä lajeilla, kuten *C. intestinalis* –lajilla, on mielenkiintoinen tutkimuskohde evoluutiobiologian kannalta. Lisäksi se voi tarjota tietoa kokonaisvaltaiseen oreksiinisysteemin ymmärtämiseen myös ihmisillä.

Tämä väitöskirja on toteutettu yhteistyössä Helsingin Yliopiston Farmaseuttisen Kemian ja Teknologian osaston (Farmasian Tiedekunta) ja Eläinlääketieteellisten Biotieteiden osaston (Eläinlääketieteellinen Tiedekunta) kanssa.

Acknowledgements

Finally, having this book in my hand, I want to thank all the people and organisations that have contributed in the making of this book and the journey behind it. First of all, I want to thank my supervisors. Dr. Henri Xhaard, thank you for giving me the opportunity to join your group, originally as a research assistant. Little I knew when I started a three month summer job with a bunch of sequences. I admire your expertise and perfectionism: you have always pushed me to do a little better. Prof. Jyrki Kukkonen, thank you for teaching me so much about wet lab and orexin receptors. You truly are an orexin and cell signalling specialist, and I have enjoyed working in your lab (except with that cAMP assay), you have also supported and prepared me to be an independent scientist.

I'm grateful for my custos and thesis committee member Prof. Jari Yli-Kauhaluoma, for being supportive and always so helpful person around. I also want to thank my other thesis committee member Dr. Jacqueline Moustakas-Verho for following and guiding my PhD studies.

I thank Dr. Xavier Deupi i Corral for willing to participate my dissertation as an opponent. I also wish to thank my pre-examiners Dr. Tiina Salminen and Dr. Juha Savinainen for giving me useful and kind feedback.

I want to show my gratitude also my collaborators Dr. Mikko Airavaara and Dr. Chris Langmead. Mikko, I admire your positive attitude, and I have enjoyed my time in your lab and with your nice lab members. Chris, thank you for giving me an opportunity to get to know you, your world class lab in Monash University and your friendly colleagues.

I will also thank my co-authors and collaborators, all GLORIA consortium and 3i Regeneration project members. Special thanks goes to Mrs. Tiina Ahonen for being excellent peer support in scientific challenges.

I also want to thank my closest co-workers, current and former members of CDD research group. Dr. Leo Ghemtio, Mr. Evgeni Grazhdankin, Mr. Ashenafi Legehar, Mr. Loïc Dreano, Mr. Mael Briand, Ms. Céline Perrot, Mr. Lauri Urvas, Dr. Lasse Karhu, Dr. Ainoleena Turku, Dr. Vigneshwari Subramanian, Dr. Alexandre Borrel, Dr. YueZhou Zhang, Dr. Gloria Wissel and Dr. Aniket Magarkar, thank you for cheering up my days with nice discussions, breaks, trips, events. My thanks goes also to lab technicians, especially to Mr. Santeri Suokas, who has made my working days in lab smooth, and to Mr. Kaj-Roger Hurme for helping me with isotopes.

I am grateful for Waldemar von Frenckell, Oskar Öflund and Emil Aaltonen foundations, GLORIA and 3i Regeneration projects for funding my research.

I will thank Helsinki Roller Derby, especially lovely All-Stars, my sisters from another misters. Thank you for many years of training, games, friendship and being there for me, and thank you for my nicknames Tiede-Maiju and Bilsa-Muija. I enjoy to identify all the plants and insects you find in your ways. I also want to thank my biologist friends “Käpylehmät” Tiina, Piksa, Martta and Henna (you are counted to this list) and your spouses. Your interest in science admires me, and your support has meant a world to me.

Mom (and dad too), I want to thank you for all the support and encouragement during my studies. Your interest in nature has led my way to this point, and I am thankful for all the opportunities you have made possible to me. I also want to thank my twin sister Varpu and big sister Tytti for always being there for me, and also for sharing a horrible sense of humour with me.

Finally, I will thank my partner Tiina. Thank you for your love and believing in me, you are the most supportive and kind person I know. Thank you for lowering my stress levels when needed by bringing me to forest with our lovely dogs Rölli and Piku (thanks to them too!). And thank you for your genuine interest in very different area of science what you do.

Table of Contents

List of original publications

This dissertation is based on the following publications referred to in the text by the Roman numerals I–IV.

I **Rinne M**, Tanoli ZU, Khan A, Xhaard H. Cartography of rhodopsin-like G protein-coupled receptors across vertebrate genomes. Sci Rep 9:7058, 2019.

II **Rinne MK**, Leino TO, Turku A, Turunen PM, Steynen Y, Xhaard H, Wallén EAA, Kukkonen JP. Pharmacological characterisation of the orexin/hypocretin receptor agonist Nag 26. Eur J Pharmacol 837:137-144, 2018.

III Turku A, **Rinne MK**, Boije Af Gennäs G, Xhaard H, Lindholm D, Kukkonen JP. Orexin receptor agonist Yan 7874 is a weak agonist of orexin/hypocretin receptors and shows orexin receptor-independent cytotoxicity. PLoS One 12:e0178526, 2017.

IV **Rinne MK**, Fridmanis D, Riddy D, Langmead CJ, Kukkonen JP, Xhaard H. Pairing a peptide ligand to a *Ciona intestinalis* putative orexin receptor. *Manuscript*

Personal contributions

I Author was responsible for data retrieval, sequence analysis, manual annotation, phylogenetic reconstruction and manuscript writing. H. Xhaard planned and supervised the study and contributed to manuscript writing. Z.U. Tanoli was responsible for automated data management. A. Khan contributed by co-supervising the study.

II Author contributed to study design together with J.P. Kukkonen and conducted the majority of the experiments and data analysis. Author contributed to the manuscript writing together with J.P. Kukkonen, A. Turku and H. Xhaard. T.O. Leino, Y. Steynen and E. Wallén were responsible for the synthesis.

III Author conducted part of the experiments and data analysis with J.P. Kukkonen and A. Turku. J.P. Kukkonen and A. Turku were responsible for the study design. Author participated in manuscript writing together with H. Xhaard, A. Turku and J.P. Kukkonen.

IV Author was responsible for study design together with H. Xhaard and J.P. Kukkonen and for designing and conducting molecular cloning. Author conducted the majority of the experiments and writing the manuscript together with H. Xhaard and J.P. Kukkonen.

Additional publications

V Ahonen TJ, **Rinne M**, Grutschreiber P, Mätlik K, Airavaara M, Schaarschmidt D, Lang H, Reiss D, Xhaard H, Gaveriaux-Ruff C, Yli-Kauhaluoma J, Moreira VM. Synthesis of 7β-hydroxy-8-ketone opioid derivatives with antagonist activity at mu- and delta-opioid receptors. Eur J Med Chem. 151:495-507, 2018.

VI **Rinne MK**, Mätlik K, Ahonen TJ, Vedovi F, Zappia G, Moreira VM, Yli-Kauhaluoma J, Leino S, Salminen O, Airavaara M, Xhaard H. Topoisomerase II inhibitor Mitoxantrone and two analogues blocking Toll-like receptor 4 mediated activation of NFκB. *Manuscript*

Abbreviations

%ID	Percentage of sequence identity
1R	First round of whole-genome duplications
1-SORA	OX_1-selective antagonist
2R	Second round of whole-genome duplications
2-SORA	OX_2-selective antagonist
3D	Three-dimensional
3R	Third round of whole-genome duplications
AA	Amino acid(s)
AB	Antibody
AC	Adenylyl cyclase
ATP	Adenosine triphosphate
BBB	Blood–brain barrier
BLAST	Basic local alignment search tool
cAMP	Cyclic adenosine monophosphate
CHO	Chinese hamster ovary
Ci-orexin-A	Putative orexin peptide of *C. intestinalis*
CiOX	Putative orexin receptor of *C. intestinalis*
CiPPO	*C. intestinalis* prepro-orexin
CNS	Central nervous system
CRE	cAMP response element
CTx	Cholera toxin
DAG	Diacylglycerol
DORA	Dual orexin receptor antagonist
EC_{50}	Half-maximal effective concentration
ECL	Extracellular loop
ELK-1	ETS Like-1 protein
E_{max}	Maximum effect
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
FDA	Fluorescein diacetate
FDA	Food and Drug Administration
FRT	Flippase recognition target
$GABA_A$	Gamma-aminobutyric acid A receptor
GAPs	GTPase-activating proteins

GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GPCR	G protein-coupled receptor
GRK	GPCR kinase
GTP	Guanosine triphosphate
GTPase	Guanosine triphosphatase
HBM	HEPES-buffered medium
HEK	Human embryonic kidney
HTS	High throughput screening
ICL	Intracellular loop
IP_3	Inositol 1,4,5-triphosphate
K_i	Inhibitory constant
LGR	Leucine-rich repeat receptors
MAPK	Mitogen-activated protein kinase
MCH	Melanin-concentrating hormone receptors
MECA	Melanocortin/EDG/cannabinoid/adenosine receptors
MRG	Mas-related receptors
NCBI	National Center for Biotechnology Information
NFAT	Nuclear factor of activated T-cells (response element)
NMR	Nuclear magnetic resonance
NSCC	Non-selective ion channel
ORF	Open reading frame
OX_1	(Human) OX_1 orexin receptor
OX_2	(Human) OX_2 orexin receptor
PBS	Phosphate-buffered saline
PDB	Protein Data Bank code
PEI	Polyethyleneimine
PI	Propidium iodide
PIP_2	Phosphatidylinositol 4,5-bisphosphate
PKA	Protein kinase A
PKC	Protein kinase C
PLC	Phospholipase C
PLD	Phospholipase D
PMDA	Pharmaceuticals and Medical Devices Agency

PTx	Pertussis toxin
REM	Rapid eye movement
RGS	Regulator of G-protein signalling
RhoGEF	RhoGTPase nucleotide exchange factors
SAR	Structure-activity relationship
SEM	Standard error of the mean
SOC	Store-operated calcium channel
SOG	Somatostatin/opioid/galanin receptors
SOREM	Sleep-onset REM
SRE	Serum response element
TAMRA	5- and 6-carboxytetramethylrhodamin
TM	Transmembrane
TRPC	Transient receptor potential-canonical channel
WT	Wild-type

1 Introduction

G protein-coupled receptors (GPCRs) are integral membrane proteins sharing a canonical 3-dimensional structure of seven transmembrane α-helical segments (TM1–7) connected by intra- and extracellular loops (Bockaert and Pin, 1999; Palczewski et al., 2000; Clark, 2013; Wang et al., 2013) and were named for their ability to activate (or inhibit) second messenger pathways through the activation of heterotrimeric G proteins (Lefkowitz, 1994). They constitute one of the largest families of proteins in humans (Lander et al., 2001), with about 800 members, approximately 400 of which are olfactory receptors (Malnic et al., 2004). GPCRs take part in various physiological processes such as control of blood pressure, anxiety, pain responses, immune system, smell, sight and taste (e.g., Sutherland, 1971; Burbach, 2004). To do so, they are activated by diverse exogenous ligands, such as olfactants, and endogenous ligands, such as neuropeptides, hormones and neurotransmitters. Pharmaceutical drugs can also act as ligands and can activate or inhibit these receptors.

GPCRs, therefore, are key targets of the pharmaceutical industry; to date, about 30–50% of drugs on the market either directly or indirectly target these receptors (Hauser et al., 2017). Furthermore, an intense area of research is chemical biology, the identification of small molecules necessary to study biological functions, distribution and modes of action (Ohlmeyer and Zhou, 2010; Hughes et al., 2011). Recent years have been the golden age in GPCR research, with fascinating topics of high pharmaceutical and pharmacological relevance, such as biased signalling, G protein-binding interface, allosteric modulators, or receptor dimerisation that open many new avenues to be exploited. Many new three-dimensional (3D) structures have been solved at atomic resolution, and at the time of writing this thesis, there are 69 publically available unique structures (Irvine, 2019). Individual receptors' structures can further be stabilised and trapped along different conformational intermediates, for example, using nanobodies or different compounds binding at allosteric or orthosteric sites (Piscitelli et al., 2015; Dong et al., 2016). These 3D structures shed light on different aspects of GPCR signalling and ligand interaction but also on their evolution (Wolf and Grünewald, 2015).

Yet, the large repertoire of GPCRs together with the difficulty of apprehending complex cellular events makes their study a daunting task. Many GPCRs were cloned at the end of the 80s or during the 90s (Nathans and Hogness, 1983; Dixon et al., 1986), while others were revealed by genome sequencing studies (Ruuskanen et al.,

2004). These later putative receptors were either demonstrated to be novel subtypes or orphan receptors (receptors with unknown ligand and function) that were progressively deorphanised; to date, about 50 orphan receptors remain to be paired with a cognate ligand in humans (Fang et al., 2015). Overall, human receptors have received the most attention, although non-human signalling systems have considerable potential, e.g., as model systems more simple than human (Bosch et al., 2017), to design animal models in toxicology studies (Segner and Baumann, 2016) or, more generally, to understand the effect of pharmaceutical compounds released in the environment (Nystén et al., 2019).

The first aim of the thesis (Publication **I**) is to investigate the repertoire of GPCRs in non-human vertebrates and early vertebrates, which is critical for planning animal models, predicting ecotoxicology or, more generally, for understanding evolutionary and biological functions. Although individual repertoires have been characterised in many invertebrate and vertebrate species (Brody and Cravchik, 2000; Lagerström et al., 2006; Gloriam et al., 2007; Kamesh et al., 2008; Sarkar et al., 2011; Nagarathnam et al., 2012; Krishnan et al., 2013), these studies did not classify and identify individual subtypes. Another major difficulty is with the nomenclature: naming new receptors often follows the human counterpart (Lagerström et al., 2006; Gloriam et al., 2007; Kamesh et al., 2008; Sarkar et al., 2011), but receptors without human counterparts (such as new subtypes) are often named in a way that is inconsistent with their evolutionary relationships, or they are not named at all (Yamamoto et al., 2013; Harding et al., 2017; Hoyer, 2017). Following Publication **I**, some material is still in preparation and is therefore not added in this thesis, but it will be published at a later time: the cartography of the receptors in Families B-F, and the identification of new receptors in non-human vertebrate species from these families. Additionally, mapping invertebrate and early vertebrate receptors to their human counterparts has been conducted.

The second aim of the thesis focuses on the orexin system, both in an evolutionary perspective by cloning and pairing a putative orexin receptor from the vase tunicate *Ciona intestinalis* (Publication **IV**) and by characterising the functional response to the compounds Nag26 and Yan7874 (Publications **II** and **III**) (Yanagisawa, 2010; Nagahara et al., 2015). Orexin receptors, also known as hypocretin receptors, belong to the rhodopsin peptide GPCRs, and two subtypes, OX_1 and OX_2 orexin receptors, exist in humans. These receptors are encoded by the HCRTR1 and HCRTR2 genes, respectively. In humans, these receptors can be

activated by two endogenous peptides named orexin-A and orexin-B (33 and 28 amino acids long, respectively), which result from the cleavage of a prepro-orexin precursor (gene name HCRT). The orexin neuropeptides were discovered simultaneously by two different research groups, which led to two parallel nomenclatures: orexin (Sakurai et al., 1998), based on the role in appetite, and hypocretin (de Lecea et al., 1998), based on the expression in the hypothalamus. The orexin system is an essential modulator of arousal, reflected by the excitatory effect of orexins on cholinergic and monoaminergic systems in many cell types (reviewed in Leonard and Kukkonen, 2014), and therefore, it is also an important regulator of sleep and wakefulness. Orexin antagonists, widely pursued by large pharmaceutical companies, have a main indication against insomnia, while agonists have been suggested to be useful against sleep disorders such as narcolepsy but have received much less attention. The orexin system also plays a role in stress, appetite and reward. At the start of the thesis (2013), no small-molecule agonist was reported; Publication **II** characterised the signalling pathways of the potent orexin receptor agonist Nag26 (Nagahara et al., 2015), while publication **III** characterised another potential agonist, Yan7874, referenced only from a patent owned by the University of Texas (Yanagisawa, 2010). Suvorexant, a first-in-class therapeutic compound for the treatment of insomnia that acts on orexin receptors, was approved in 2014 by the Food and Drug Administration (FDA; USA) and later by the Pharmaceuticals and Medical Devices Agency (PMDA; Japan) (Jacobson et al., 2014; Kuriyama and Tabata, 2017). The last part of the thesis, publication **IV**, focused on an orphan putative orexin GPCR from a tunicate (*C. intestinalis*). The receptor was cloned and tested for its effect on human peptides, small molecules and predicted peptides. This study is still ongoing at the time of writing, and only the best results obtained have been condensed into a manuscript (Publication **IV**). The projected publication time is late 2019. Additional material, for example, attempts to characterise the pathways at Monash University using reporter assays (research group of Chris Langmead) or construction of chimeric (humanised) ascidian receptors exchanging intracellular loops have been conducted but will not be presented. A set of studies about modelling morphinean binding modes at opioid receptors and virtual screen at the toll-like receptor 4, followed by *in vitro* validation, were also produced during the thesis work (see Other Publications) but are not included in this thesis in order to preserve its focus.

This thesis thus comprises a set of computational and pharmacological studies from sequence-based receptor classification and homology modelling to

pharmacological studies of small molecules and peptides to orexin receptors. The review of the literature gives an overview of the GPCR structure and on the orexin system.

2 Review of the literature

2.1 Pharmacology of G protein-coupled receptors

G-protein coupled receptors are transmembrane proteins sharing a similar 3-dimensional structure of seven transmembrane (TM1–7) α-helices connected by intra- and extracellular loops (Palczewski et al., 2000; Wang et al., 2013). The extracellular loops (ECL1–3) and the N-terminus lie outside the cell, and the intracellular loops (ICL1–3) and the C-terminus lie inside. These structures form three domains: extracellular, transmembrane and intracellular. The α-helices form a bundle, and at the centre, towards the extracellular surface, a crevice can form the binding site of endogenous ligands, named the orthosteric site (in opposition to allosteric ligand binding sites). However, it is worth mentioning that mostly only GPCRs from the rhodopsin family have an orthosteric binding site between the TMs (see below), while the members of the other families commonly bind with N-terminus to extracellular signalling molecules (Schiöth and Fredriksson, 2005; Gacasan et al., 2017).

There are two families of models for GPCR activation, conformational selection and conformational induction, and both are compatible with existing data (Kenakin, 1996). Both models are prone to the existence of active and inactive forms (that can be stabilised and structurally solved; see below). In conformational selection, the receptor "oscillates" between conformations, some of which may be stabilised by certain ligands; for conformational induction, the conformational changes follow (and induce) ligand binding. Conformational selection is supported by constitutive activity (signalling in the absence of ligands) exhibited by some GPCRs. Activation of the receptor by ligand induction (agonist, Figure 1) or conformational stabilisation (or both) then leads to an exchange of guanosine diphosphate (GDP) to guanosine triphosphate (GTP) on an intracellular G protein, proceeding to signal transduction (Strange, 2008, reviewed in Rosenbaum et al., 2009). The G protein-coupling is enabled by a movement of TM6 that protrudes outward after ligand binding (Manglik and Kruse, 2017), which is made possible by breaking a salt bridge to TM3 in a fraction of GPCRs (the so-called ionic lock). Roughly half of the amino acid contacts of the receptor are reorganised upon activation and G protein coupling (Venkatakrishnan et al., 2016). Common reorganisation of residue contacts in cytosolic region are identified into positions 3.46, 6.37 and 7.53 (Ballesteros–Weinstein numbering; see paragraph 2.5 Methods

to study molecular evolution) in class A GPCRs despite their diverse ligand binding or G protein coupling preferences, and thus are key residues in converging the receptor activation to G protein mediated signalling. There is evidence that some GPCRs can be G protein-coupled without agonist binding, e.g., constitutively activated receptors (Seifert and Wenzel-Seifert, 2002) or pre-coupled inactive receptors (e.g., Qin et al., 2011). Conversely, inverse agonist (Figure 1) binding stabilises inactive receptor conformation(s), which then inhibits possible constitutive signalling activity of the receptor. Antagonists, in turn, do not affect the receptor conformation as such but, rather, obstruct the agonist binding.

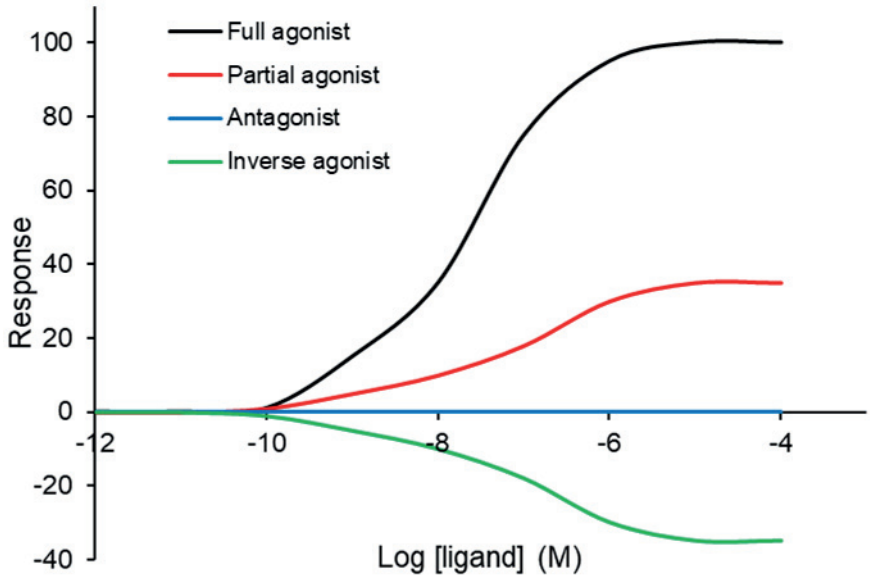

Figure 1. Exemplified concentration response curves of full agonist (maximum response, 100%), partial agonist, antagonist (neutral) and inverse agonist. The X-axis represents basal activity.

The expression of GPCRs has been studied by following the GPCR-associated protein complexes that match with the life cycle of GPCRs (reviewed in Daulat et al., 2009; Maurice et al., 2011). Post-translational modifications and folding take place in the endoplasmic reticulum (ER) with an interaction with chaperone proteins (the translocon complex) that ensure sufficient folding by masking the hydrophobic regions and unpaired cysteines and disulfide bridge formation. After folding, GPCRs are packed into vesicles and exported to the Golgi for further maturation. Mature GPCRs are transported to their destination, e.g., plasma membrane, and misprocessed

GPCRs go back to the ER for refolding or degradation. It has been suggested that possible dimerisation and oligomerisation already happen in the folding process in the ER (Bulenger et al., 2005).

2.2 Heterotrimeric G proteins

As the name suggests, GPCRs couple with G proteins, though the coupling has not been verified for all (Alexander et al., 2013). Heterotrimeric G proteins are membrane-anchored complexes formed by subunits Gα, Gβ and Gγ. Receptor signalling is mediated by Gα and/or the Gβγ complex. In total, there are 18 Gα, 5 Gβ and 12 Gγ subunit proteins found in humans (reviewed in Syrovatkina et al., 2016). In its inactive state, the Ras-like/guanosine triphosphatase (GTPase) domain of Gα is bound to GDP, which is exchanged for free intracellular GTP during the ligand-triggered receptor activation, and this further leads to the dissociation of the Gα (+ GTP) subunit and the Gβγ complex (reviewed in Gilman, 1987; Coleman et al., 2015; Syrovatkina et al., 2016). Both the Gα subunit and Gβγ complex activate separate signalling cascades, but many of the responses result from the cooperative action of Gα and Gβγ. Gα subunits possess intrinsic GTPase activity, which is able to hydrolyse GTP to GDP (reviewed in Milligan and Kostenis, 2006). GTPase activity is also regulated by several other proteins, such as GTPase-activating proteins (GAPs).

The hydrolysis of GTP to GDP – often with help from the regulator of G-protein signalling (RGS) proteins – leads the receptor back to the inactive state, and again, to the association of G-protein subunits to heterotrimeric G-protein. As a desensitisation mechanism, many of the GPCRs are phosphorylated by GPCR kinases (GRKs), followed by β-arrestin binding and receptor internalisation to endosomes, where the receptors are either recycled back to the plasma membrane or degraded (reviewed in Irannejad, 2014; Gurevich and Gurevich, 2019). β-arrestins also act as signal transducers, e.g., by activating mitogen-activated protein kinase (MAPK) pathways. Additionally, it has been suggested that β-arrestin-mediated internalisation plays a role in pathway selection (reviewed in Hoyer and Bartfai, 2012; Pavlos and Friedman, 2017). In addition to GRK-regulated phosphorylation, activated protein kinase C (PKC) and protein kinase A (PKA) regulate desensitisation by phosphorylating the receptor, which leads to the uncoupling of the receptor from the G protein. In some cases, PKA can regulate the G-protein switch, in which the

receptor couples with another G protein after phosphorylation (Daaka et al., 1997; Lawler et al., 2001).

Receptors are not always activated in a similar manner; e.g., involving the same Gα subunit(s). Activation can vary based on compound properties, and different compounds can activate different pathway(s) (reviewed in Galandrin et al., 2007). This phenomenon is called biased signalling, and it most likely depends on receptor conformation after ligand binding. This has been of interest in drug research, especially in cases where receptor activation leads to adverse effects via specific pathways.

The 18 human Gα subunits are divided into four main families, G_q, G_i, G_S and G_{12}, all having their main signalling effectors, such as various protein kinases (e.g., MAPK, PKC), calcium elevation, adenylyl cyclase (AC), phospholipases and potassium channels (reviewed in Syrovatkina et al., 2016). The G_q-activated phospholipase C (PLC) and the G_S-activated adenylyl cyclase pathways are well characterised, whereas G_i inhibits adenylyl cyclase (Figure 2), and G_{12} activates Rho signalling, among others. However, crosstalk between signal pathways is relatively common (Katoh et al., 1998; Aittaleb et al., 2010, reviewed in Sunahara et al., 1996), and defining the specific pathway from the receptors to the measured response is often challenging. Furthermore, receptors often signal through more than just one Gα subunit family (reviewed in Galandrin et al., 2007).

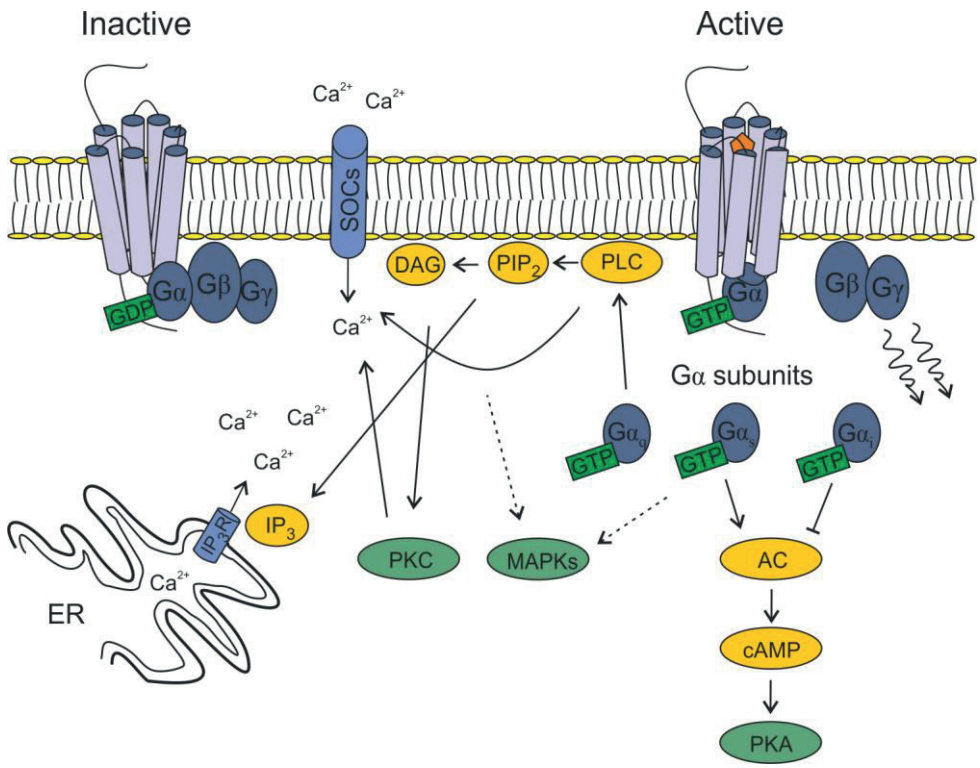

Figure 2. Schematic and simplified presentation of G protein-mediated signalling. The dissociation of Gα and Gβγ after GTP binding and three Gα subunit-mediated signalling cascades are presented. Adenylyl cyclase (AC) is located in the plasma membrane but not placed there for clarity of the figure. G_{12} or Gβγ-activated signalling are not described. Ligand (orange pentagon) binding leads to receptor activation and G protein-mediated signalling. Upon receptor activation, GTP binds to the Gα subunit, which leads to the dissociation of Gα and Gβγ. Hydrolysis of GTP to GDP leads back to an inactive state of protein and association of Gα and Gβγ. In the active state, the $G\alpha_q$ subunit activates phospholipase C (PLC), which hydrolyses phosphatidylinositol-4,5-bisphosphate (PIP_2) to diacylglycerol (DAG) and inositol-1,4,5-trisphosphate (IP_3). IP_3 then induces calcium efflux from the endoplasmic reticulum (ER) by binding to the IP_3 receptor. Calcium depletion from the ER, protein kinase C (PKC) and PLC also regulate store-operated calcium channels (SOCs) and extracellular calcium influx. $G\alpha_s$ activates AC, which leads to an elevation of intracellular cyclic adenosine monophosphate (cAMP), and further activation of protein kinase A (PKA). $G\alpha_i$, in turn, inhibits AC. Both $G\alpha_q$ and $G\alpha_s$ also indirectly activate mitogen-activated protein kinases (MAPKs).

Gα subunit Gq – G_q is responsible for the activation of PLC β (reviewed in Rhee and Bae, 1997). The activation of PLC is followed by the hydrolysis of phosphatidylinositols, such as phosphatidylinositol-4,5-bisphosphate (PIP_2), which gives rise to the secondary messengers diacylglycerol (DAG) and inositol-1,4,5-trisphosphate (IP_3). It should be noted that the PIP_2 also has an important regulatory role itself; lipophilic DAG remains in the plasma membrane, but IP_3 is released into the cytosol. DAG activates, e.g., PKC, and IP_3 induces Ca^{2+} elevation by binding its cognate receptor on the ER membrane. The intracellular calcium-elevating mechanism of PLC is two-phased. IP_3 receptors act like Ca^{2+} channels, and activation leads to the release of stored calcium from the ER into the cytosol. PLC (along with PKC and Ca^{2+} depletion in the ER) also activates store-operated calcium channels (SOCs) in the plasma membrane and results in an extracellular calcium influx, which can further activate, e.g., extracellular signal-regulated kinase 1/2 (ERK1/2).

Gα subunit G_S and G_i – G_S and G_i regulate the activity of AC (reviewed in Sunahara et al., 1996). G_S acts as an AC activator and, conversely, G_i as an inhibitor of AC. G_S-promoted activation of AC leads to an increase in intracellular cyclic adenosine monophosphate (cAMP), a derivative of ATP, and further to the activation of cAMP-activated PKA. G_S is also indirectly linked to MAPK activation. Besides the inhibitory effect of G_i on AC, the Gβγ complex from G_i can directly enhance G_q-promoted PLC activation, and it can activate potassium channels and activate or inhibit AC (depending on the form of AC) independently from G_i (Tang and Gilman, 1991; Sunahara et al., 1996). Additionally, the Gβγ complex from G_i indirectly promotes MAPK activation.

Gα subunit G_{12} – G_{12} is the smallest and least well-known family of heterotrimeric G proteins. G_{12} promotes the activation of RhoGEFs, members of the small G protein superfamily, that further stimulate, e.g., the sodium–hydrogen exchange (Suzuki et al., 2009). Also, G_q promotes Rho activation (Aittaleb et al., 2010). G_{12} is also an indirect activator of phospholipase D (PLD) and PKC, but the mechanisms are not fully understood.

2.3 Neuropeptides and neuropeptide receptors

2.3.1 Neuropeptides

Neuropeptides and neurohormones are signalling molecules that vary in size; the shortest peptides are only a few amino acids long, and the longest ones are up to 100 amino acids. Neuropeptides are used in communication between neurons as well as in autocrine and paracrine agents in many organs (Burbach and Meijer, 1992). Neuropeptides are secreted by neurons, but they target various peripheral tissue in addition to the central nervous system (CNS). They participate in many physiological functions such as memory, learning, arousal and appetite, as well as behaviour, and they can modulate long-range hormonal and short-range synaptic or paracrine transmission.

Neuropeptides are encoded by propeptide genes that often contain more than one active neuropeptide (Burbach and Meijer, 1992). Prepropeptides also include sites for enzyme-targeted cleavage and signal peptide, responsible for post-translational localisation of the precursors (Steiner, 1998). The secondary structure of many neuropeptides is built around an α-helical 3D structure (Figure 3). Neuropeptides are primarily produced as inactive precursors and are further modified post-translationally to active neuropeptides (cleavage, C-terminal amidation etc.) in intracellular vesicles (reviewed in Eipper et al., 1992; Hook et al., 2008). Neuropeptides are secreted by exocytosis, often together with neurotransmitters (reviewed in Hoyer and Bartfai, 2012). Secretion results from neuronal bursting activity that often follows from physiological stimulation such as stress or pain. Neuropeptides mostly target neuropeptide receptors in the rhodopsin family included in the peptide (β-branch) and somatostatin/opioid/galanin (SOG) (γ-branch) receptors. Activation of the receptors modulates the downstream effectors in cellular systems, resulting in changes at the organ level. The discovery of small-molecule neuropeptide mimics has been challenging and slow, even if this branch of chemistry – peptidomimetics – has received considerable attention in the last decades (reviewed in Hoyer and Bartfai, 2012).

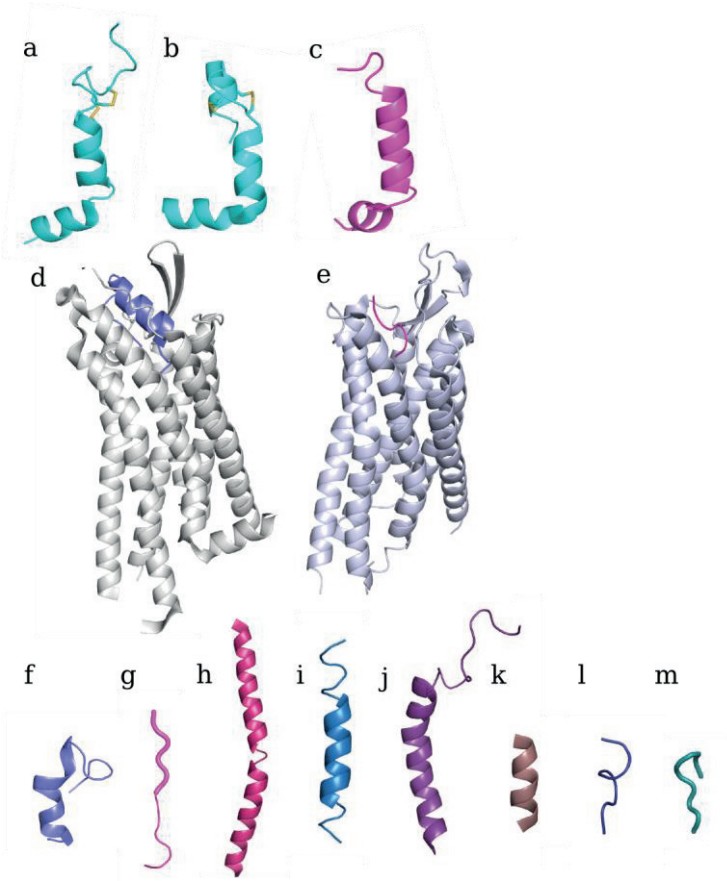

Figure 3. Three-dimensional structure of human orexin peptides, two different solution structures of orexin-A (a; PDB: 1R02 (Kim et al., 2004) and b; 1WSO (Takai et al., 2006)), and a solution structure of orexin-B (c; PDB: 1CQ0) (Lee et al., 1999). Complexes of available neuropeptide receptors with co-crystallised neuropeptides (d; endothelin receptor - endothelin complex; PDB: 5GLH (Shihoya et al., 2016), e; neurotensin receptor type 1 - neurotensin complex 4GRV (White et al., 2012)). Structures of neuropeptides targeting rhodopsin-like neuropeptide GPCRs (f-n), endothelin (f; PDB: 1EDN) (Janes et al., 1994), neurotensin (g; PDB: 2OYV) (Coutant et al., 2007), neuropeptide K (h; PDB: 2B19)(Dike and Cowsik, 2006), tachykinin neuropeptide gamma (i; PDB: 2MCE) (Chandrashekar et al., 2004), neuropeptide Y (j; PDB: 1RON) (Monks et al., 1996), neurokinin B (k; PDB: 1P9F) (Mantha et al., 2004), cholecystokinin-8 (l; PDB: 1D6G) (Pellegrini and Mierke, 1999), oxytocin (n; PDB: 2MGO) (Koehbach et al., 2013). Sequences below.

2.3.2 Orexin neuropeptides

In humans, orexin peptides are synthesised in the hypothalamus (Sakurai et al., 1998) by a relatively small number of orexinergic neurons (Peyron et al., 1998; Thannickal et al., 2000). Additionally, prepro-orexin mRNA has been detected in several tissues in the periphery (Jöhren et al., 2001). However, the knowledge of the functionality of the orexin system in the periphery is scarce. Despite the low number of orexin-producing neurons, they project to several targets in the CNS; e.g., the exciting effect on many cholinergic and monoaminergic neurons has been studied (reviewed in Leonard and Kukkonen, 2014). Orexin stimulation induces depolarisation of neurons, initiates firing or increases the level of firing, and thus, it is an important regulator of neuroexcitation (reviewed in Kukkonen and Leonard, 2014). The non-selective ion channels (NSCCs) provoke depolarisation by the Na^+/Ca^{2+} exchange and potassium channel inhibition.

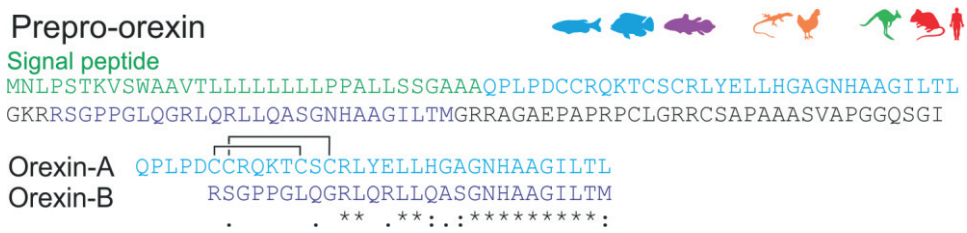

Figure 4. Human orexin precursor and diversity of orexin propeptide through species groups (absent in amphibian, platypus, invertebrates and lamprey) in the Ensembl database. Amino acid sequence of the human orexin precursor (top) and the two mature peptides orexin-A and orexin-B (bottom). Disulphide bridges (C6-12, C7-C14) in orexin-A are indicated by lines; C-terminal amidation not indicated. Star indicates corresponding amino acid, colon conservative amino acid substitution; dot indicates semi-conservative substitution; and space indicates non-conservative substitution.

The 131-amino acid-long prepro-orexin-encoded endogenous ligands, orexin-A and orexin-B, are 33 and 28 amino acids long, respectively, and share 46% sequence identity (%ID) (Figure 4). Structures of orexin-A and orexin-B have been studied in aqueous solution using nuclear magnetic resonance (NMR), with two distinct conformations of orexin-A (Protein Data Bank codes [PDB]: 1R02, 1WSO) (Kim et al., 2004; Takai et al., 2006) and one conformation of orexin-B (PDB: 1CQ0) (Lee et al., 1999) (Figure 3). The affinity of the endogenous peptides for orexin

receptor subtypes varies: orexin-A reaches a 10-fold affinity to OX_1 compared to OX_2, while orexin-B possesses a similar affinity to both subtypes (Sakurai et al., 1998); however, the selectivity is often expression system–dependent (Putula et al., 2011a). In a simplistic view, orexin-A possesses two characteristic sides, hydrophobic and hydrophilic (Miskolzie and Kotovych, 2003; Takai et al., 2006; German et al., 2013). Orexin-A contains four cysteine residues that form disulfide bonds (C6–C12 and C7–C14) (Sakurai et al., 1998). The reduction of these disulfide bridges results in a decrease in the activity (Okumura et al., 2001). Also, several truncated and point-mutated orexin-A and orexin-B peptides have been studied to determine the essential amino acids and structure-activity relationship (SAR) (Darker et al., 2001; Lang et al., 2004; Takai et al., 2006; German et al., 2013). The C-terminal fragment has been identified to be essential for ligand binding and receptor activation, while the hydrophobic N-terminal takes part in recognition. N-terminal truncations lead to a drop in activity, and the shortest active peptides in the nanomolar range have been described: the C-terminal fragment of 19-amino acid-long orexin-A (amino acids (aa) 15-33) (or even aa 17-33 (German et al., 2013)), and the C-terminal fragment of 19-amino acid-long orexin-B (aa 10-28) (Darker et al., 2001; Lang et al., 2004). Additionally, the selectivity of modified peptides have been studied; e.g., fragmental and two-amino acid-substituted [Ala^{11}, d-Leu^{15}] orexin-B show selectivity (23–400-fold) for OX_2 over OX_1 (Asahi et al., 2003; Putula et al., 2011a). The point mutation of orexin-A by alanine substitution has shown that, the closer the C-terminus of the substituted amino acid is, the more negatively it affects the activation (German et al., 2013). In fact, substituting any of the last five amino acids of the C-terminus (aa 29-33) results in a practically inactive peptide. Interestingly, modifications of orexin-A seem to be better tolerated by OX_2 than by OX_1 (Okumura et al., 2001; Putula et al., 2011b; German et al., 2013).

2.3.3 Neuropeptide receptors

Neuropeptide receptors share the similar 7-TM topology as other rhodopsin GPCRs. These receptors belong to the peptide subfamily that comprises, for example, orexin receptors, endothelin-related receptors, neurotensin receptors, neuropeptide Y receptors and cholecystokinin receptors. Neuropeptides also activate GPCRs outside the peptide subfamily; galanin receptors and opioid receptors, members of the SOG family, are endogenously activated by neuropeptides, galanin receptors by galanin

and opioid receptors by endorphins, enkephalins and dynorphins. In fact, most of the SOG and melanin-concentrating hormone (MCH) receptors bind peptide ligands. Despite the similar ligand binding (peptide) preferences, structural similarities between the receptors are missing (Fredriksson et al., 2003). Outside the rhodopsin family, many peptide GPCRs are also found, for example, calcitonin or frizzled. Many CNS-mediated pathophysiological conditions, such as insomnia, depression, and neuropathic pain, are associated with malfunctions of the neuropeptide or its cognate receptor. All conditions mentioned are in high need of better therapeutics.

Neuropeptide receptors are mostly expressed in the CNS but also in the peripheral nervous system. Expression levels tend to be relatively low; consequently, these receptors are interesting drug targets with a beneficially low number of receptors to be targeted (reviewed in Hoyer and Bartfai, 2012). Additionally, often only a nanomolar concentration of endogenous ligands is needed to achieve a full response of neuropeptide receptors, and chronic agonist treatment results only in moderate desensitisation. For example, gonadotropin-releasing hormone receptors possess no intracellular C-terminus; they are incapable of binding β-arrestin and are therefore slow to be internalised (Pawson et al., 2008). However, it is worth mentioning that the internalisation is not always the mechanism of desensitisation (reviewed in Hoyer and Bartfai, 2012; Pavlos and Friedman, 2017). As these receptors are endogenously activated by neuropeptides, their binding sites are also typically large. In drug discovery, targeting these receptors has often been directed towards a peptidomimetic approach, and the discovery of small-molecule drugs has been challenging and slow (reviewed in Hoyer and Bartfai, 2012).

The X-ray crystallography structure of several neuropeptide receptors has been published with co-crystallised small-molecule antagonists or agonists, e.g., orexin receptors (PDBs: OX_1: 4ZJ8 (Yin et al., 2016a), 4ZJC (Yin et al., 2016a), OX_2: 4S0V (Yin et al., 2015), 5WQC (Suno et al., 2018), 5WS3 (Suno et al., 2018)) from the peptide subfamily, and, e.g., opioid receptors (PDBs: μ: 4DKL (Manglik et al., 2012), 5C1M (Huang et al., 2015), κ: 4DJH (Wu et al., 2012), δ: 4EJ4 (Granier et al., 2012)) from the SOG subfamily. Endogenous peptide-bound complexes are also available for some neuropeptide receptors: e.g., endothelin receptor with co-crystallised endothelin (5GLH (Shihoya et al., 2016)) and neurotensin receptor type 1 with co-crystallised neurotensin (4GRV (White et al., 2012)), however fully active conformation was not reached in case of neurotensin receptor (Deupi, 2014).

In most GPCRs, the ECL2 forms a disulfide bridge to TM3 via conserved cysteine residue (Ruuskanen et al., 2004), and in many cases, removing the disulfide bridge has a crucial effect on membrane localisation and agonist binding (Davidson et al., 1994; Noda et al., 1994; Zeng et al., 1999; Wheatley et al., 2012). In most rhodopsin-like receptors, ECL2 is the longest of the loops, with an exception of the short-looped melanocortin receptors (Holst and Schwartz, 2003). Neuropeptide GPCRs of the A family usually share the same fold of their ECL2: the ECL2 of the β_2-adrenoceptor is an α-helical structure, whereas the ECL2 of, e.g., endothelin, neurotensin (Figure 3), orexin and rhodopsin receptors comprises two β-hairpin-swirled β-sheets that, together with the N-termini, form a lid on the TM-bundle (Wheatley et al., 2012; Yin et al., 2016b).

2.3.4 Human orexin receptors

The expression pattern of the orexin receptor subtypes in the CNS is wide and partly overlapping, but it differs especially in sleep–wakefulness-related brain areas (Kilduff and Peyron, 2000; Marcus et al., 2001). Both subtypes are also found in the periphery, e.g., in the pituitary and adrenal glands, the gastrointestinal tract and the pancreas (Jöhren et al., 2001; Zhang et al., 2005). Expression patterns differ, e.g., in the testis, the kidney and the thyroid gland where only OX_1 has been identified, and in the lungs where only OX_2 has been identified (Jöhren et al., 2001).

Peptide sequences of OX_1 and OX_2 receptors are 425 and 444 amino acids long, respectively. The sequences share 64 %ID, and if only TMs are compared, the %ID increases to 80%. X-ray structures of inactive conformations with co-crystallised antagonists are resolved for both subtypes (PDBs: OX_1: 4ZJ8 (Yin et al., 2016a), 4ZJC (Yin et al., 2016a), OX_2: 4S0V (Yin et al., 2015), 5WQC (Suno et al., 2018), 5WS3 (Suno et al., 2018)). Additionally, mutagenesis studies have clarified antagonist binding (Malherbe et al., 2010; Tran et al., 2011; Heifetz et al., 2012; Putula and Kukkonen, 2012). The active conformation is not yet solved, but several site-directed mutagenesis studies (see, e.g., Putula et al., 2011b; Heifetz et al., 2012; Nicole et al., 2015) and molecular dynamics simulations (Karhu et al., 2015) give information about the functional activity and the orthosteric binding site of the agonists (Putula et al., 2011b; Tran et al., 2011). The orexin receptor subtypes share a similar backbone of 3D structure (Yin et al., 2016b). The orthosteric binding site lies mainly within TMs 3, 5, 6 and 7, and the binding pockets of human OX_1 and OX_2

vary most importantly in the TM2 position 2.61 (T111 to S103, respectively) and the TM3 position 3.33 (T135 to A127, respectively). TM3, in particular Q3.32, has been described to be critical for small-molecule binding (common in many rhodopsin GPCRs), while interactions relevant for orexin peptides are distributed in a larger area of the receptors: TM1, TM3, TM5 and N-terminus (Tran et al., 2011). The X-ray structures, however, have shown few direct polar contacts between protein and (antagonist) ligands, while most contacts are water-mediated (Yin et al., 2016b). The N-termini of OX_1 and OX_2 form a short α-helical segment with an extension towards TM1 in OX_1, which is essential for interaction with orexin-A (Yin et al., 2016b). A shared structure in ECL2 forms a structure with β-hairpin-connected β-sheets, contains a conservative cysteine bridge with TM3, and is important for ligand binding and receptor activation (Malherbe et al., 2010).

2.3.5 Orexin receptor signalling cascades

Orexin receptor signalling and G protein coupling is well described (reviewed in Kukkonen and Leonard, 2014; Leonard and Kukkonen, 2014), and signalling at the cellular level has been described in many recombinant cell systems as well as in native cells (see, e.g., Sakurai et al., 1998; Lund et al., 2000; Karteris et al., 2001; Randeva et al., 2001; Milasta et al., 2005). The G protein coupling of orexin receptors can vary between subtypes and between different tissues (Karteris et al., 2001, 2005; Randeva et al., 2001). Altogether, signalling mechanisms of orexin receptors are versatile and complex.

Orexin receptors signal mainly through G_q, leading to intracellular calcium elevation (see Figure 2). Calcium elevation is mostly a consequence of the activation of PLC and further IP_3, which activates the IP_3 receptors and calcium release from the ER. PLC, Ca^{2+} depletion (in the ER) and PKC -regulated activation of ion channels, such as NSCCs, transient receptor potential-canonical channels (TRPCs), and SOCs, in the plasma membrane, leads to an extracellular calcium influx (Kukkonen and Åkerman, 2001; Peltonen et al., 2009). TRPC activation and extracellular calcium influx are not dependent on IP_3 (Ekholm et al., 2007), contrary to intracellular calcium release. Orexin receptor–mediated PLC activation also activates PKC (Xia et al., 2009; Jäntti et al., 2012). Also, other phospholipases such as phospholipase A_2 and D have been shown to be involved in orexin signalling cascades (Johansson et al., 2008; Jäntti et al., 2012; Turunen et al., 2012). The G_q

signalling pathway (and PLC activation) also activates p38 MAPK and ERK1/2, and the ERK activation has been shown to be Ca^{2+} influx-dependent (Milasta et al., 2005; Ammoun et al., 2006). Further, EKR1/2 activation can also be modulated by PKA, AC activity and β-arrestin, while calcium elevation and PKC can activate AC (reviewed in Kukkonen and Leonard, 2014).

Orexin receptors couple and signal additionally through G_S and G_i families and also interact with other proteins such as β-arrestin. By coupling to G_S and G_i, orexin receptor signalling also activates and inhibits, respectively, AC and further cAMP and PKA (Malendowicz et al., 1999; Randeva et al., 2001; Karteris et al., 2005; Woldan-Tambor et al., 2011; Urbańska et al., 2012). Orexin-A-induced signalling prefers G_i (and G_q) over G_S, that is, activated only by a high concentration of orexin-A (reviewed in Leonard and Kukkonen, 2014).

Activation of orexin receptors can lead to the binding of β-arrestin. As for many other GPCRs, the binding of β-arrestin is associated with receptor desensitisation and is often followed by receptor internalisation by endocytosis (Ward et al., 2011). Additionally, β-arrestin acts as a signalling protein by activating p38 MAPK and ERK1/2 (Milasta et al., 2005).

2.3.6 Small-molecule modulators of orexin receptors, and therapeutic opportunities

2.3.6.1 Pathophysiology of the orexin system

Several pathophysiological conditions are associated with the orexin system. As a strong regulator of sleep and wakefulness, a malfunctioning orexin system may cause insomnia or narcolepsy. It has been suggested that in insomnia, the disorder of promoting or maintaining sleep, the orexin system remains activated, e.g., because of the overexpression of orexinergic neurons and/or high levels of orexins. Intracerebroventricular orexin-A administration in rat models promotes arousal, wakefulness and activity, as well as neuronal firing activity in many brain regions (reviewed in Brisbare-Roch et al., 2007). Typically, in sleep periods, orexin-A levels in the cerebrospinal fluid are about half of the levels in waking and active periods, being lowest after a continuous sleep period and highest after a long, continuous active period (Yoshida et al., 2001, reviewed in Brisbare-Roch et al., 2007). Pharmacologically blocking orexin receptors promotes both non-rapid eye movement

and rapid eye movement (REM) sleep, the latter of which is difficult to achieve with traditional gamma-aminobutyric acid A receptor ($GABA_A$)–targeted pharmaceutics. In contrast to insomnia, symptoms of narcolepsy are often daytime somnolence, as well as cataplexy, sudden loss of muscle tone thought to originate from the loss of orexinergic neurons, and/or low levels of orexin-A (Siegel and Boehmer, 2006). Disruption of the orexin system, by, e.g., OX_2 receptor mutation or knockout, prepro-orexin knockout, or genetic destruction of the orexinergic neurons, has been shown to lead to narcolepsy-like phenotypes in canine and mice models, suggesting a central role of the orexin system in narcolepsy (Chemelli et al., 1999; Lin et al., 1999; Willie et al., 2003; Beuckmann et al., 2004). Additionally, OX_1 has a weaker role in sleep-wakefulness regulation, and knockout of OX_1 does not lead to a narcolepsy-like phenotype in rodents (Mieda et al., 2011).

Orexin neuropeptides were originally identified as regulators of appetite (Sakurai et al., 1998). Orexin signalling promotes feeding behaviour by activating the production of hunger- and feeding-stimulating neuromodulators, such as neuropeptide Y, in neurons (Edwards et al., 1999). Orexin signalling has a direct effect on hunger and food intake, but it is also evident in general energy homeostasis (review in Girault et al., 2012). Short-term orexin stimulation leads to higher food intake, but not necessarily to an increase in body weight. In turn, deficient orexin production, e.g., in orexin knockout mice, may lead to obesity and insulin resistance (Tsuneki et al., 2008). On the contrary, long-term orexin stimulation prevents the development of obesity and insulin resistance and enhances leptin sensitivity (Funato et al., 2009). However, long-term orexin stimulation had no effect in leptin knockout mice. OX_2 receptor signalling has a stronger role in feeding-related disorders, while OX_1 knockouts do not show an apparent effect on body mass but instead have a protective effect on insulin levels (Funato et al., 2009).

Reward-seeking behaviour is associated with the development and maintenance of many addictions. The role of orexin signalling in reward-seeking has been studied mostly in the mesocorticolimbic reward pathway, where orexin receptors are highly expressed in many areas (reviewed in Sharf et al., 2010). Partially, the link between orexins and the reward system arises from the direct effect of orexins on the dopaminergic neurons in the ventral tegmental area, and thus, orexins have an evident role in drug-induced neurotransmission and further synaptic plasticity (Korotkova et al., 2003; Baimel and Borgland, 2012). Several drug-specific responses in drug abuse have been described in the lateral hypothalamus and in the

orexin system. Short-term abuse of, e.g., cocaine or amphetamine often leads to an increased expression of the transcription factor c-Fos in the orexinergic neurons, while long-term abuse of, e.g., nicotine leads to a higher expression of orexin receptors and orexin peptides. In addition to the effect of drugs, the c-Fos level in the orexinergic neurons has been shown to depend on the behavioural state, such as sleep and wake states (Estabrooke et al., 2001). The link between drug-seeking behaviour and the orexin system has also been studied. Blocking orexin receptors, particularly OX_1 with SB-334867 (see 2.3.6.3 Small-molecule antagonists), decreased alcohol and nicotine intake (Lawrence et al., 2006; Hollander et al., 2008), while orexin-A administration into the lateral hypothalamus increased alcohol intake (Schneider et al., 2007). However, the pattern is not consistent for all drugs (reviewed in Sharf et al., 2010).

The orexin system is also associated with acute and chronic stress responses (reviewed in Sargin, 2018). Intracerebroventricular orexin administration leads to typical stress-related behaviour as well as to an increase in stress hormones in plasma and neuronal c-Fos levels in the paraventricular nucleus of the hypothalamus in rodents. Additionally, orexin neurons are activated in acute anxiety, panic and fear, and provoked responses can be blocked with an OX_1 antagonist but not with OX_2 antagonists.

Additional pathophysiological conditions of the orexin system are nociception, in which antinociceptive effects are mediated mostly by activation of OX_1 (Holland and Goadsby, 2007; Ho et al., 2011), and depression, in which OX_1 activation promotes depressive behaviour, while OX_2 activation promotes antidepressive behaviour (reviewed in Summers et al., 2018). Both conditions are associated with the level of wakefulness and stress. A rather surprising relation of the orexin system to cancer is described in the periphery, e.g., gastrointestinal tract (Rouet-Benzineb et al., 2004). Exposure of OX_1-expressing colon cancer cells to orexins leads to OX_1 activation-dependent apoptosis.

2.3.6.2 Small-molecule agonists

Similarly to other neuropeptide receptors, orexin receptor–targeting small-molecule agonist development has proceeded slowly (for summary, see Table 1) (reviewed in Hoyer and Bartfai, 2012). So far, only one orexin receptor–targeting series of compounds is described with a nanomolar range of half-maximal effective

concentration (EC_{50}) values on OX_2 but with low micromolar EC_{50} values on OX_1. Nag26 (Figure 5), the most potent compound of the series, showed EC_{50}=23 nM on OX_2 and EC_{50}=1616 nM on OX_1 for Ca^{2+} mobilisation in the recombinant Chinese hamster ovary cell line (CHO) (Nagahara et al., 2015). All active compounds of the series are OX_2-selective, with Nag26 showing a 70-fold selectivity for OX_2 according to that study. Nag26 was developed from a sulfonamide group–containing hit compound from high throughput screening (HTS). The sulfonamide group is essential for the activity of Nag26; replacing the group with carbonylamide led to an inactive compound. Due to the low water solubility of Nag26 compound 30 (YNT-185, Figure 5 (Irukayama-Tomobe et al., 2017)) was developed by replacing the dimethyl substituent of the B-ring (Figure 5) with a dimethylamino group. YNT-185 is highly soluble in water, and it shows EC_{50}-values of 28 nM on OX_2 and 2750 nM on OX_1 in the Ca^{2+} mobilisation assay in recombinant CHO cells, being ~100-fold more potent on OX_2 (Irukayama-Tomobe et al., 2017). Stimulation with YNT-185 has been described to start firing of histaminergic neurons in the tuberomammillary nucleus of the hypothalamus *in vitro*, and intracerebroventricular administration increases wakefulness in wild-type (WT) mice without affecting the body temperature or heart rate, and it decreases SOREM (sleep-onset REM, cataplexy-like) state in narcolepsy-like prepro-orexin knockout mice, but not in orexin receptor knockout mice. However, YNT-185 is not lipophilic enough to cross the blood–brain barrier (BBB).

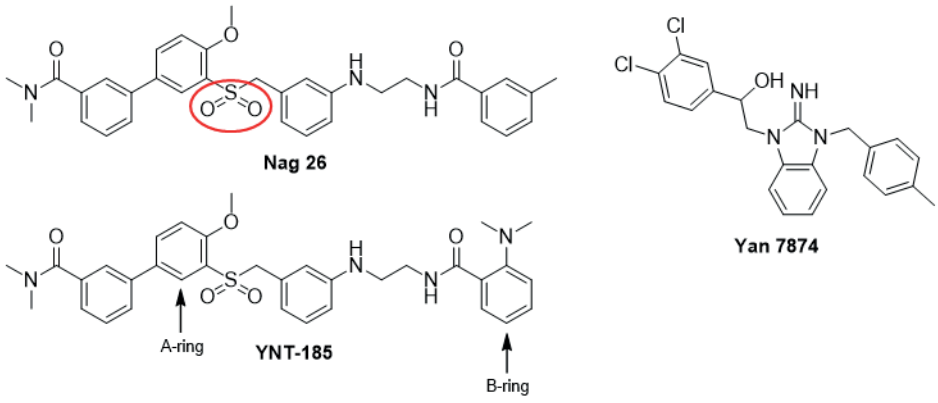

Figure 5. Small-molecule orexin receptor agonists Nag26, YNT-185 and Yan7874. Sulfonamide group, crucial for binding of Nag26 (and YNT-185), circled with red. B-ring modification of Nag26 led to YNT-185 and improved water solubility but to a slight drop in activation in OX_2 (EC_{50} 23 nM → 28 nM), and rather high drop in OX_1 (EC_{50} 1616 nM → 2750 nM).

A few other series of OX_2-selective agonists have been reported through patents, such as the lead compound of the Yanagisawa patent, Yan7874 (Figure 5 (Yanagisawa, 2010)), and the lead compound of the Cano patent, compound 28 (Cano et al., 2014), but there is still very limited pharmacological data. Yan7874 showed increased Ca^{2+} levels in OX_2-expressing recombinant CHO cells, while in OX_1-expressing cells, Ca^{2+} elevation was lower. Additionally, Yan7874 was reported to activate the nuclear factor of activated T-cells (NFAT) and cAMP response element (CRE) luciferase in OX_2-expressing CHO cells co-transfected with reporter genes. NFAT luciferase increased clearly (fold induction contradictory in the patent) from basal at 10 μM concentration. *In vivo*, orally administered Yan7874 showed a suppression of the narcolepsy-like state and promoted wakefulness in orexin neuron–deficient mice. Activation of the compounds of the Cano patent (Cano et al., 2014) was reported with an intracellular Ca^{2+} increase at a 50 μM concentration in the OX_2-expressing recombinant undisclosed cell line. The lead compound, compound 28, gave a response of 93% of the orexin-induced maximum effect (E_{max}).

Also, two compounds from the series of azulene-based compounds have been reported to show weak OX_2 agonism (Leino et al., 2018). The same series also introduces orexin receptor potentiators (potentiating the response of known agonists) and antagonists. At 10 μM, concentration compounds 7 and 27 showed 6% and 5% of orexin-A induced E_{max} on OX_1, and 12% and 11% on OX_2, respectively, in a calcium elevation study. When an unspecific effect was ruled out, compounds showed 5% of E_{max} induced by orexin-A only on OX_2. SARs of compounds have been described in the subsequent publication (compounds 1 and 2) (Turku et al., 2019). Another publication by Turku and co-workers in 2016 (Turku et al., 2016) describes the identification of other weak agonists by pharmacophore-based virtual screening and further characterisation with Ca^{2+} elevation. Compounds (4-7) showed slight selectivity for OX_1 at 10 μM concentration, showing 2.7–7.3% of orexin-A induced E_{max}, while OX_2 activation was 1.0–3.7%.

Table 1. Summary of orexin receptor agonists with EC_{50} and E_{max} values on both orexin receptor subtypes. N/A indicates data not available.

Agonist	Method	Cell line	EC_{50} (nM) OX_1	EC_{50} (nM) OX_2	E_{max} OX_1	E_{max} OX_2	IUPAC name
Nag26	Ca^{2+}	CHO	1616	23	100 %	98 %	4'-methoxy-N,N-dimethyl-3'-[N-(3-{[2-(3-methylbenzamido)ethyl]amino}phenyl sulfamoyl]-(1,1'-biphenyl)-3-carboxamide
YNT-185	Ca^{2+}	CHO	2750	28	55 %	94 %	2-(dimethylamino)-N-[2-[3-[[5-[3-(dimethylcarbamoyl)phenyl]-2-methoxyphenyl]sulfonylamino]anilino]ethyl]benzamide
Yan7874	Ca^{2+}, NFAT, CRE	CHO	N/A	N/A	N/A	N/A	1-(3,4-dichlorophenyl)-2-[2-imino-3-(4-methylbenzyl)-2,3-dihydro-1H-benzo[d]imidazol-1-yl]ethan-1-ol
Cano 28	Ca^{2+}	N/A	N/A	N/A	93 %	N/A	2-(2-amino-3,6-dimethylphenoxy)-3-chloronaphthalene-1 ,4-dione
Leino 7	Ca^{2+}	CHO	N/A	N/A	0 %	5 %	Methyl 1-[4-(trifluoromethyl)benzoyl]azulene-6-carboxylate
Leino 27	Ca^{2+}	CHO	N/A	N/A	0 %	5 %	Methyl 2-(1-benzoylazulen-6-yl)acetate
Turku 4	Ca^{2+}	CHO	N/A	N/A	6 %	1 %	1-(9H-carbazol-4-yloxy)-3-[2-(2-methoxyphenoxy)ethylamino]propan-2-ol
Turku 5	Ca^{2+}	CHO	N/A	N/A	7 %	3 %	Methyl 2-[acetyl-[[5-bromo-2-[(2-chlorobenzoyl)amino]phenyl]-phenylmethyl]amino]acetate
Turku 6	Ca^{2+}	CHO	N/A	N/A	5 %	2 %	6-(3,5-dimethylpyrazol-1-yl)-4-N-phenyl-2-N-[(E)-[(E)-4-phenylbut-3-en-2-ylidene]amino]-1,3,5-triazine-2,4-diamine
Turku 7	Ca^{2+}	CHO	N/A	N/A	3 %	4 %	2,6-dibromo-4-[(Z)-[[4-(furan-2-ylmethylamino)-6-(4-methoxyanilino)-1,3,5-triazin-2-yl]hydrazinylidene]methyl]phenol

2.3.6.3 Small-molecule antagonists

Unlike only a few described orexin receptor agonists, several orexin-targeting small-molecule antagonists have been reported (for summary, see Table 2), mainly for the treatment of insomnia. Orexin receptor antagonists are divided into three groups: OX_1-selective antagonists (1-SORAs), OX_2-selective antagonists (2-SORAs) and dual orexin receptor antagonists (DORAs) (Figure 6). However, only one compound, the DORA suvorexant (Figure 6, originally known as MK-4305) is approved by the FDA (USA) and PMDA (Japan) for the treatment of insomnia (Kuriyama and Tabata, 2017).

Diazepane-cored suvorexant was developed by Merck in a series of several DORAs (Cox et al., 2010), and being the most promising compound with a high affinity with OX_1- and OX_2-inhibitory constant (K_i) values of 0.55 nM and 0.35 nM, respectively, it was successfully assessed *in vivo* and further in clinical trials. Suvorexant has a low off-target effect, and it promotes non-REM and REM sleep throughout vertebrate species (Winrow et al., 2011). Suvorexant has also been successfully crystallised in a complex with OX_2 (PDB: 4RNB) (Yin et al., 2015). The other Merck compound from the same series, the piperidine-amide-cored filorexant (MK-6096), also a potent DORA, has also reached clinical trials. *In vitro* filorexant binds to OX_1 and OX_2 with affinities (K_i) of 2.5 nM for OX_1 and 0.31 nM for OX_2, and it has been shown to have more potential than suvorexant in preclinical experiments by fast occupancy of OX_2 receptors at low plasma concentrations (Winrow et al., 2012, reviewed in Janto et al., 2018). Another piperidine-amide-cored compound, SB-649868, also reached clinical trials, but those were later discontinued due to preclinically seen toxicity (Cox et al., 2010). SB-649868 is DORA with $K_i <$ 1 nM for both orexin receptor subtypes, and it promotes both non-REM and REM sleep, and maintains sleep (Di Fabio et al., 2011, reviewed in Roecker et al., 2016).

Lemborexant, the cyclopropane-cored, high-affinity DORA with K_i values of 4.8 nM on OX_1 and 0.61 nM on OX_2 is approved as a New Drug Application (FDA) for treating insomnia (2019). Lemborexant promotes sleep, particularly non-REM and with no effect on REM in rodents (Yoshida et al., 2014). It has similar effects on humans, improving sleep efficiency with only weak adverse effects (somnolence) at low concentrations (Murphy et al., 2017). Almorexant (Figure 6, ACT-078573), a tetrahydroisoquinoline derivative, a high-affinity DORA, shows affinity values (K_i) of 7.0 nM for OX_1 and 2.6 nM for OX_2 (Beuckmann et al., 2017). Almorexant shows high selectivity for orexin receptors and an ability to cross BBB, thus decreasing

waking and increasing REM and non-REM efficiency. Consequently, it also reached clinical trials for treating insomnia, but later trials were halted due to high liver enzyme production (reviewed in Roecker et al., 2016).

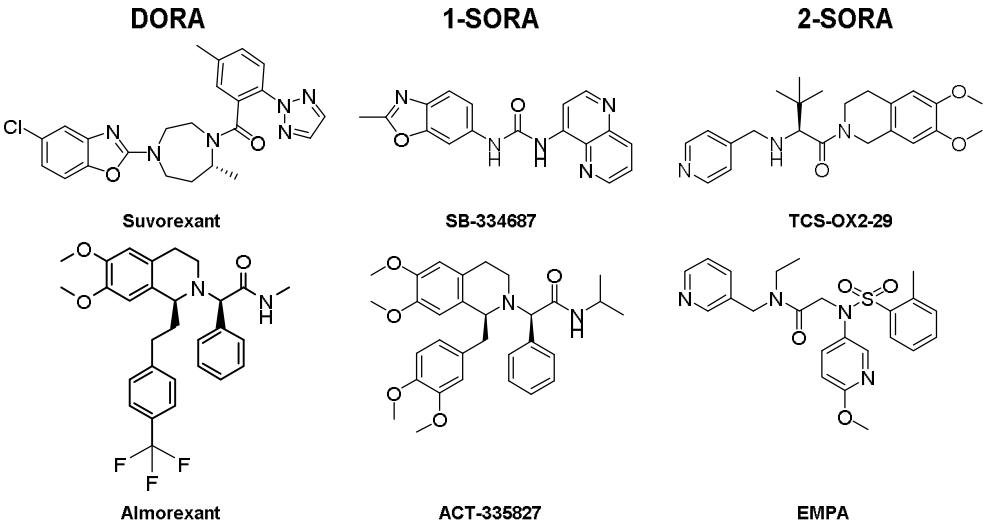

Figure 6. Examples of small-molecule orexin antagonists. DORAs suvorexant and almorexant show selectivity for both orexin receptor subtypes. 1-SORAs SB-334687 and ACT-335827 show a 117-fold and 70-fold, respectively, selectivity for OX_1. 2-SORAs TCS-OX2-29 shows no selectivity for OX_1 and EMPA shows an 818-fold selectivity for OX_2.

The therapeutic opportunities of 2-SORAs and DORAs are mostly in the treatment of insomnia, while the opportunities of 1-SORAs are less understood. It has been suggested that 1-SORAs could have potential in addiction and anxiety pharmacotherapeutics. SB-334867 (Figure 6) was described as the first OX_1-selective antagonist (Smart et al., 2001). With the K_i values of 18 nM and 835 nM on OX_1 and OX_2, respectively, SB-334867 shows a 46-fold selectivity for OX_1 (reviewed in Stump et al., 2016). SB-334867 is widely used as a pharmacological tool, but with relatively poor selectivity (orexin receptors in general, and OX_1 among subtypes), hydrolytic stability and bioavailability, which limits its use *in vivo* (McElhinny et al., 2012; Stump et al., 2016). However, SB-334867 treatment has been shown to decrease nicotine- and alcohol-seeking in rodents (Lawrence et al., 2006; Hollander et al., 2008). Another compound with better bioavailability, 1-SORA ACT-335827 (Figure 6), with K_b-values of 41 nM on OX_1 and 560 nM on OX_2, a 14–70-fold

selectivity for OX_1, depending on compared values (Steiner et al., 2013). ACT-335827 reduces anxiety-like behaviour and fear in rodents, without a decrease in wakefulness.

Other urea-cored compounds described as 1-SORAs, SB-674042, SB-408124 and SB-410220 all show low/mid nanomolar (1.1–22 nM) affinity in calcium mobilisation assay for OX_1, with 117-, 65- and 58-fold selectivity for OX_1, respectively (Langmead et al., 2004). SB-674042 has shown a preventive effect on long-term depression (Park and Weon, 2017). Moreover, compounds with a morphinan skeleton and a sulfonamide-containing side chain in tertiary amine have been described as 1-SORAs (Nagase et al., 2017), with a K_i-value of 1.36 nM on OX_1 for the lead compound 71. No inhibitory effect was seen on OX_2.

2-SORAs have therapeutic opportunities in sleep disorders but also in eating disorders. Tetrahydroisoquinole (Figure 6, TCS OX2 29) was the first published 2-SORA, with an IC_{50}-value of 40 nM on OX_2 and no reported inhibition of OX_1 binding/response at a concentration up to 10 μM (Hirose et al., 2003). More pharmacological data are found for the 2-SORA JNJ-10397049 (compound 9): K_i-values on OX_1 and OX_2 are 1644 nM and 6 nM, respectively, giving a 274-fold selectivity for OX_2 (McAtee et al., 2004). JNJ-10397049 has been reported to promote sleep in rodents, but because of its poor drug-like properties, its development has discontinued (Dugovic et al., 2009). Another 2-SORA from the same research group, JNJ-42847922, shows an 80-fold selectivity for OX_2, with K_i-values of 800 nM and 10 nM on OX_1 and OX_2, respectively (Letavic et al., 2015). This compound has better drug-like properties but rather poor bioavailability (Bonaventure et al., 2015). JNJ-42847922 has reached clinical trials, showing a significant increase in somnolence.

The 2-SORA EMPA (Figure 6) shows an 818-fold selectivity for OX_2 with 900 nM and 1.1 nM K_i-values on OX_1 and OX_2, respectively (Malherbe et al., 2009). This compound also showed effects *in vivo*, but its uptake level after oral or intravenous administration was only moderate in animal models. The only 2-SORA with *in vivo* effects similar to DORAs is MK-1064, which shows K_i-values of 1584 nM and 0.5 nM on OX_1 and OX_2 receptors, respectively, and a 3000-fold selectivity for OX_2. It increases both NREM and REM sleep in animal models without leading to narcolepsy-like state (Roecker et al., 2014, reviewed in Gotter et al., 2016).

Table 2. Summary of orexin receptor antagonists with K_i values on both orexin receptor subtypes and selectivity type (DORA, 1-SORA, 2-SORA). Binding indicates radioligand-displacement binding assay, HEK indicates HEK-239, N/A indicates data not available.

			K_i			
Antagonists	**Method**	**Cell line**	**OX_1**	**OX_2**	**Type**	**IUPAC name**
Suvorexant	Binding	CHO	0.55	0.35	DORA	[(7R)-4-(5-chloro-1,3-benzoxazol-2-yl)-7-methyl-1,4-diazepan-1-yl]-[5-methyl-2-(triazol-2-yl)phenyl]methanone
Filorexant	Binding	CHO	2.5	0.31	DORA	[(2R,5R)-5-[(5-fluoropyridin-2-yl)oxymethyl]-2-methylpiperidin-1-yl]-(5-methyl-2-pyrimidin-2-ylphenyl)methanone
SB-649868	Binding	CHO	> 1	> 1	DORA	N-{[(2S)-1-{[5-(4-Fluorophenyl)-2-methyl-1,3-thiazol-4-yl]carbonyl}-2-piperidinyl]methyl}-1-benzofuran-4-carboxamide
Lemborexant	Ca^{2+}	HEK	4.8	0.61	DORA	(1R,2S)-2-[(2,4-dimethylpyrimidin-5-yl)oxymethyl]-2-(3-fluorophenyl)-N-(5-fluoropyridin-2-yl)cyclopropane-1-carboxamide
Almorexant	Ca^{2+}	HEK	7.0	2.6	DORA	(2R)-2-[(1S)-6,7-dimethoxy-1-[2-[4-(trifluoromethyl)phenyl]ethyl]-3,4-dihydro-1H-isoquinolin-2-yl]-N-methyl-2-phenylacetamide
SB-334867	Ca^{2+}	CHO	18	835	1-SORA	1-(2-Methyl-1,3-benzoxazol-6-yl)-3-(1,5-naphthyridin-4-yl)urea
ACT-335827	IP_1	CHO	41 (K_b)	560 (K_b)	1-SORA	(2R)-2-[(1S)-1-(3,4-Dimethoxybenzyl)-6,7-dimethoxy-3,4-dihydro-2(1H)-isoquinolinyl]-N-isopropyl-2-phenylacetamide
SB-674042	Ca^{2+}	CHO	1.1 (K_b)	129 (K_b)	1-SORA	[5-(2-Fluorophenyl)-2-methyl-1,3-thiazol-4-yl]{(2S)-2-[(5-phenyl-1,3,4-oxadiazol-2-yl)methyl]-1-pyrrolidinyl}methanone
SB-408124	Binding	CHO	26.9	N/A	1-SORA	1-(6,8-Difluoro-2-methyl-4-quinolinyl)-3-[4-(dimethylamino)phenyl]urea
SB-410220	Binding	CHO	4.5	N/A	1-SORA	1-(5,8-Difluoro-4-quinolinyl)-3-[4-(dimethylamino)phenyl]urea
Nagase 71	Ca^{2+}	CHO	1.36	>10 000	1-SORA	(E)-N-((4R,4aS,7R,7aR,12bS)-3-{[2-(Dimethylamino)phenyl]sulfonyl}-4a-hydroxy-9-methoxy-2,3,4,4a,5,6,7,7a-octahydro-1H-4,12-methanobenzofuro[3,2-e]isoquinolin-7-yl)-N-methyl-3-(pyridin-2-yl)acrylamide
TCS OX2 29	Ca^{2+}	CHO	>10 000	40 (IC_{50})	2-SORA	(2S)-1-(6,7-dimethoxy-3,4-dihydro-1H-isoquinolin-2-yl)-3,3-dimethyl-2-(pyridin-4-ylmethylamino)butan-1-one
JNJ-10397049	Binding	CHO	1644	6	2-SORA	1-(2,4-dibromophenyl)-3-[(4S,5S)-2,2-dimethyl-4-phenyl-1,3-dioxan-5-yl]urea
JNJ-42847922	Binding	HEK	800	10	2-SORA	[2-(4,6-dimethylpyrimidin-2-yl)-1,3,3a,4,6,6a-hexahydropyrrolo[3,4-c]pyrrol-5-yl]-[2-fluoro-6-(triazol-2-yl)phenyl]methanone
EMPA	Binding	HEK	900	1.1	2-SORA	N-ethyl-2-[(6-methoxypyridin-3-yl)-(2-methylphenyl)sulfonylamino]-N-(pyridin-3-ylmethyl)acetamide
MK-1064	Binding	CHO	1584	0.5	2-SORA	5-(5-chloropyridin-3-yl)-N-[(5,6-dimethoxypyridin-2-yl)methyl]-2-pyridin-2-ylpyridine-3-carboxamide

2.3.7 Model organisms understanding the origin of the orexin system

Model species are often essential in fully understanding the complex signalling systems, e.g., the orexin system in humans. The orexin system has been studied, e.g., in *Danio rerio*, the zebrafish (Kaslin et al., 2004), and *Branchiostoma floridae*, the lancelet (Wang et al., 2019). In zebrafish, the orexin system resembles the human orexin system, with hypothalamic secretion and wide projections to cholinergic and aminergic neurons. The system likely participates in energy homeostasis and sleep-wakefulness regulation. The lancelet also possesses a functional orexin system. The expression of lancelet orexin was observed in a cerebral vesicle that resembles the vertebrate brain. The level of orexin was affected by temperature or feeding, and the orexin system likely regulates energy homeostasis in the lancelet. The connection to the endocrine system remained unsolved.

Another model organism, *C. intestinalis*, the vase tunicate, is an invertebrate in the phylum *Chordata* and subphylum *Tunicata* and is considered the closest living relative of vertebrates (Figure 7). Larval *C. intestinalis* possess similar structures as vertebrates, despite a remarkably lower overall cell number (reviewed in Satoh, 2003). In the larval stage, *C. intestinalis* has a dorsal neural tube and a notochord, an early spine-like structure. The well-characterised structure of the larval stage resembles a tadpole, an aquatic larval stage of amphibians, though it is less complex. An adult *C. intestinalis* possesses no nerve cord or notochord. *C. intestinalis* offers an interesting model in developmental neuroscience, genomics and evolutionary biology by sharing structures of vertebrates and invertebrates and by bridging the gap between the ancient and the more developed systems. With approximately 16 000 genes and no genome duplications (Dehal et al., 2002), *C. intestinalis* offers a great model for investigating ancient genes, and it could possibly also offer a great model for neuropeptide research. Its nervous system has been studied in order to understand, e.g., brain asymmetries and connectomes (Ryan et al., 2016) or neural regeneration (Dahlberg et al., 2009).

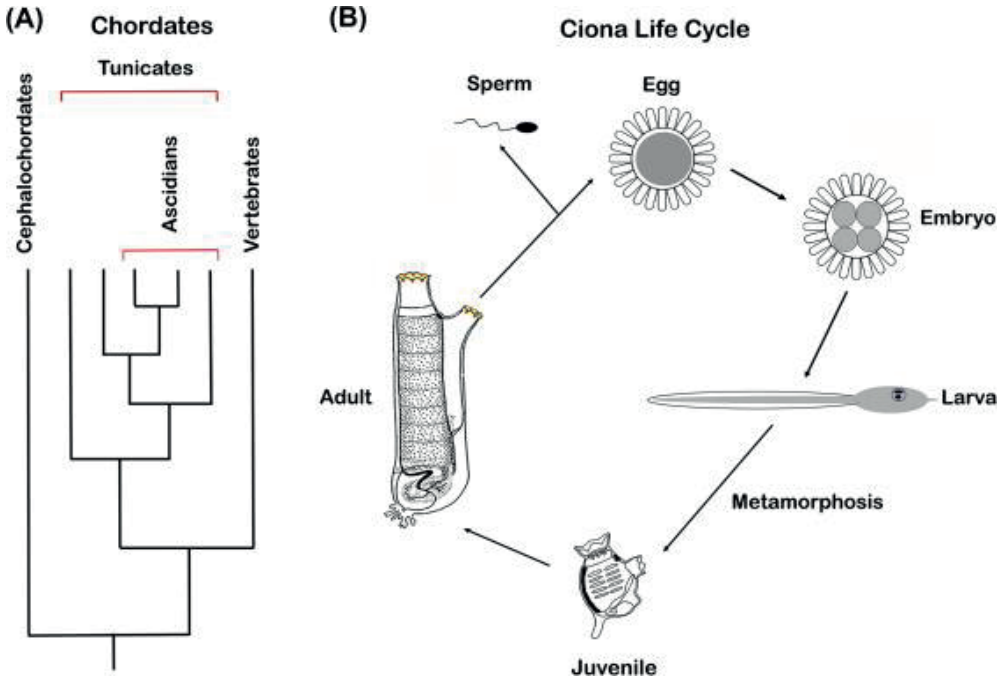

Figure 7. *Phylogenetic tree of phylum Chordata (A). The life cycle of the Ciona species includes gametic, embryonic, larval, juvenile and adult phases (B). Figure adapted from (Jeffery, 2018).*

The repertoire of the GPCRs of *C. intestinalis* has been described. The *C. intestinalis* genome was the seventh published whole genome (Dehal et al., 2002) and has been estimated to have 169 GPCRs (Kamesh et al., 2008), which represent all families of the GRAFS classification system and 1.1% of the total gene number. Similarly to humans, most GPCRs (68%) belong to the rhodopsin family.

A putative orexin receptor of *C. intestinalis* has been suggested (Kamesh et al., 2008) and has also been found in the Ensembl database as an unannotated gene (gene code ENSCING00000007467, protein code ENSCINP00000015323) clustering together with orexin receptors in the root of the phylogenetic tree. The *C. intestinalis* putative orexin receptor shares a 33% sequence identity with human orexin receptors; when the termini and ICL2 are removed, %ID increases to 38% (Figure 8). The endogenous ligand of this receptor (the putative orexin peptide) has not been

identified in *C. intestinalis*, nor are any other biological data of the receptor available, and thus, the receptor can be considered orphan.

Interestingly, in behavioural experiments, larval *C. intestinalis* exhibit sensory arousal and thigmotaxis (movement away/towards the stimuli) (Rudolf et al., 2019). Thigmotaxis can be modulated with the anxiotropic drug modafinil, which is used in the pharmacotherapy of human narcolepsy, and that has been shown to enhance the activation of orexin neurons in mice (Chemelli et al., 1999). Modafinil shows specificity for wakefulness-related brain areas, and it decreases daytime somnolence and increases wakefulness (reviewed in Salerno et al., 2019). The action mechanism of modafinil is not fully understood, and it has affinity on several binding sites. It increases Fos-immunoreactivity in orexin neurons, indicating that the wakefulness-promoting effect is possibly mediated by neuropeptide. However, modafinil increases arousal more in orexin knockouts than in wild-type mice (Willie et al., 2005), and it shows wakefulness-promoting activity even in OX_2 knockout mice (Wisor et al., 2001). Thus, any direct conclusion of the connection between the modafinil-induced arousal and orexin system in *C. intestinalis* cannot be drawn.

Considering a sequenced genome, identifying short and distant (from query sequence) neuropeptides from databases is extremely challenging (Grimmelikhuijzen and Hauser, 2012; Jekely, 2013; Mirabeau and Joly, 2013). The *C. intestinalis* neuropeptides are only rarely clustered together with vertebrate neuropeptides in partly automated databases; e.g., the Ensembl database automatically predicts only the *C. intestinalis* gonadotropin-releasing hormone that clusters together with vertebrate gonadotropin-releasing hormone (Zerbino et al., 2018). However, several *C. intestinalis* neuropeptides and peptide hormones have been described, as identified by mass spectroscopy (Kawada et al., 2011). These include, in addition to the gonadotropin-releasing hormone, at least tachykinin, oxytocin, vasopressin, calcitonin, insulin, corticotropin-releasing factor and cholecystokinin/gastrin (reviewed in Kawada et al., 2010). Functional and structural studies of the identified/suggested neuropeptides and their receptors indicate an evident link between tunicates and vertebrates, and the origin of many vertebrate neuropeptides in the tunicate lineage. These findings might have a key role in better understanding the origin but also the function of many complex signalling systems in humans, including the orexin system.

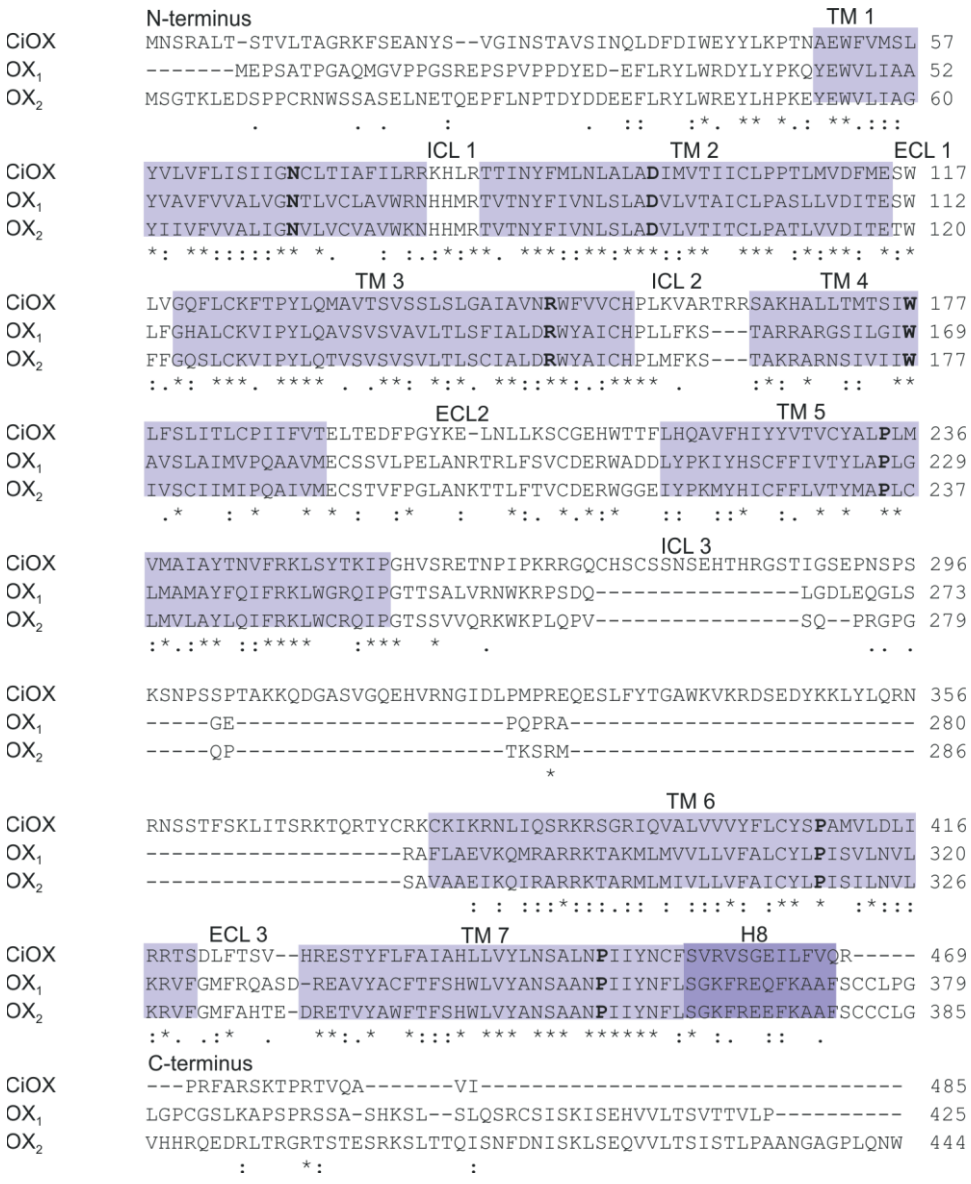

Figure 8. The aligment of human orexin receptor subtypes and putative C. intestinalis orexin receptor. α-helical (TM) regions highlighted in color, H8 presenting the additional helix of orexin receptors. Conserved pivot positions in bold. Stars indicate corresponding amino acids, colon conservative amino acid substitution, dots semi-conservative substitution, and spaces non-conservative substitution.

2.4 Evolution and classification of GPCRs

During the last 50 years or so, from the first sequencing of DNA (1977) to the present day, the methods and rapidity of sequencing have improved remarkably. In the 1990s, the first automated methods were developed, leading to an explosion in the number of sequenced genomes. The human genome, sequenced in 2000, was an important milestone (Lander et al., 2001). Following sequencing, large efforts are further necessary to assemble and compare genomic maps and to predict transcripts. The number of annotated genomes at the National Center for Biotechnology Information (NCBI) is 537, and the genomes from about 150 species are available in the Ensembl database (http://ensemblgenomes.org/, (Zerbino et al., 2018)). Additionally, sequenced genomes have been stored, e.g., in GenBank, NCBI (https://www.ncbi.nlm.nih.gov/genome/). Specialised data are stored in databases such as FlyBase (*Drosophila*, http://flybase.org/) or WormBase (*Nematode*, https://wormbase.org).

A new era in genome sequencing has come with the advent of high-throughput sequencing. It allows the sequencing, for example, of nearly identical food pathogens or of a cohort of patients. A project is now ongoing to sequence all 66 000 known animals, birds, fish and plants in the UK over the next 10 years (Wellcome Sanger Institute, 2019).

The theory of whole-genome duplications (see below) can be utilised in the prediction of approximate genome size (1R/2R/3R) (see Table 3), but not always in the prediction of the total gene number (Spring, 1997; Gregory, 2001; Garcia-Fernàndez, 2005; Steinke et al., 2006; Putnam et al., 2008). However, the largest genome to date, the axolotl, contains 10 times more bases than humans (but nearly as many genes) (Nowoshilow et al., 2018).

Table 3. Taxonomic order of species, genome sizes and the predicted gene numbers of the example species†. 1 Mb = 10^6 base pairs.*

Superphylum	Phylum	Subphylum	Example species	Genome size	Gene number
Deuterostomia	Chordata	Vertebrata	Mammals (*Homo sapiens**), birds, reptiles, amphibians, fishes (*Takifugu rubripes*†)	3,2 Gb*, 390 Mb†	19 000 – 22 000*[a], 31 000†[b]
		Tunicata	Sea squirts, *C. intestinalis**	153 – 159 Mb	16 000[c]
		Cephalochordata	Lancelets, *B. floridae**	520 Mb	21 900[d]
	Ambulacraria	Echinodermata	Starfishes, sea urchins, *Ophionotus victoriae, Strongylocentrotus purpuratus**	814 Mb	23 300[e]
		Hemichordata	Acorn worm: *Saccoglossus kowalevskii**	758 Mb	34 200[f]
Protostomia	Ecdysozoa	Arthropoda	Insects, crustaceans, etc. *Drosophila melanogaster**	165 Mb	13 600[g]
		Nematoda	*Caenorhabditis elegans**	100 Mb	19 000[h]
	Spiralia	Mollusca	Cephalopods, gastropods, *Lottia gigantea*	348 Mb	23 800[i]
		Annelida	Ringed worms, *Capitella teleta*, Helobdella robusta*†	324 Mb*, 228Mb†	32 000*, 23 000†[i]

[a] Southan, 2004; Gregory et al., 2006; Abascal et al., 2018, [b] Aparicio et al., 2002, [c] Dehal et al., 2002, [d] Putnam et al., 2008, [e] Sea Urchin Genome Sequencing Consortium et al., 2006, [f] Simakov et al., 2015, [g] Adams et al., 2000, [h] Consortium*, 1998, [i] Simakov et al., 2013

Many GPCRs were discovered after the completion of the human genome project in the year 2000. Estimates of the vertebrate GPCR repertoires suggest hundreds of functional gene products: 799 in human (Gloriam et al., 2007), 557 in chicken (Lagerström et al., 2006), 1867 and 1783 in rat and mouse, respectively (Gloriam et al., 2007), 316 in pufferfish (Sarkar et al., 2011), and 169 and 260 in early vertebrates, the tunicate *C. intestinalis* (Kamesh et al., 2008) and the Florida lancelet *B. floridae* (Krishnan et al., 2013), respectively. GPCRs are also found in invertebrate species; for example, they number approximately 1100 in the roundworm *Caenorhabditis elegans* (Bargmann, 1998) and approximately 300 in the fruit fly *Drosophila melanogaster* (Brody and Cravchik, 2000). GPCRs are also found in plants (Plakidou-Dymock et al., 1998), yeast *Saccharomyces cerevisiae* (Dohlman et al., 1991) and slime mold *Dictyostelium discoideum* (Devreotes, 1994), as well as in protozoa (Vernier et al., 1995; New and Wong, 1998). With regard to GPCRs, invertebrates tend to have a more narrow gene repertoire, possessing only one representative gene compared to the four/eight gene map of vertebrates. However, several factors affect the genome size and gene number, such as chromosomal fusions/fissions, as well as intrachromosomal changes, impact gene density and secondary deletions and duplications.

2.5 Methods to study molecular evolution

This thesis aims to classify GPCR sequences into new subtypes and receptor groups. Here, the wording classification implies determining a reasonable hypothesis about the evolutionary events that allow the suggestion of either new subtypes or new orphan receptors. The most common events in GPCR evolution are, for example, genome or large-block duplications, gene duplications and deletions (turning into pseudogenes); the addition of domains (gene fusion) has occurred in the glutamate family (for example, the Venus flytrap domain (O'Hara et al., 1993)), but the rhodopsin family is exempt for the most part. Establishing evolutionary hypotheses usually require a combination of experimental (pharmacological characterisation, gene expression) and computational techniques (analysis of genomic data, sequences, or more rarely, structures). This is usually done for a set of closely related clades: species- or receptor subtype group–specific annotations have been made, for example, for the neuropeptide Y, dopamine, pyroglutamylated RFamide peptide, relaxin family peptide, endothelin receptors and α_2 adrenoceptors (Fredriksson et al.,

2004; Xhaard et al., 2006; Hyndman et al., 2009; Good et al., 2012; Yegorov and Good, 2012; Yamamoto et al., 2013; Larhammar et al., 2014).

Phylogenetic studies – Molecular phylogenetics is the branch of phylogeny that analyses protein sequences to determine evolutionary events. In order to do so, sequences are aligned and equivalent positions are considered independent evolutionary sites. Alignment is often done at the protein level due to the added reliability in aligning sequences from 20 amino acids instead of 4 nucleotides (Opperdoes and Lemey, 2009). In the rhodopsin family, each TM contains conserved amino acids, which are mostly responsible for signal transduction, "a recognition and connection region". The most conserved motifs of each helix are well characterised, and they are often used in the identification of the putative GPCRs. The Class A GPCR-conserved motifs in each helix are as follows: TM1: GN, TM2: L/AxxD, TM3: E/DRY, TM4: W, TM5: FxxP, CWxP, and TM7: NPxxY (reviewed in Nygaard et al., 2009). For each motif, the most conserved amino acid, the pivot position, has been determined (Figure 9, Table 2) (Isberg et al., 2014), and they can be used to guide the GPCR sequence alignments. The Ballesteros–Weinstein numbering is based on the scheme of the pivotal amino acids; the numbering follows the TM number (the first digit) and distance (in relation to the number of amino acids) from the pivot position (second digit) (Ballesteros and Weinstein, 1995). The pivot position of each helix is numbered at 50, and the numbering ascends towards the C-terminus.

Table 2. Conserved amino acid motifs of each transmembrane helix in rhodopsin-like receptors. Pivotal amino acids are underlined. Percentages indicate the conservation of the pivot amino acid in human rhodopsin-like receptors (Isberg et al., 2014).

TM	Motifs	Conservation of pivot
TM1	G<u>N</u>	98%
TM2	L/Axx<u>D</u>	90%
TM3	E/D<u>R</u>Y	95%
TM4	<u>W</u>	97%
TM5	Fxx<u>P</u>	78%
TM6	CWx<u>P</u>	99%
TM7	N<u>P</u>xxY	88%

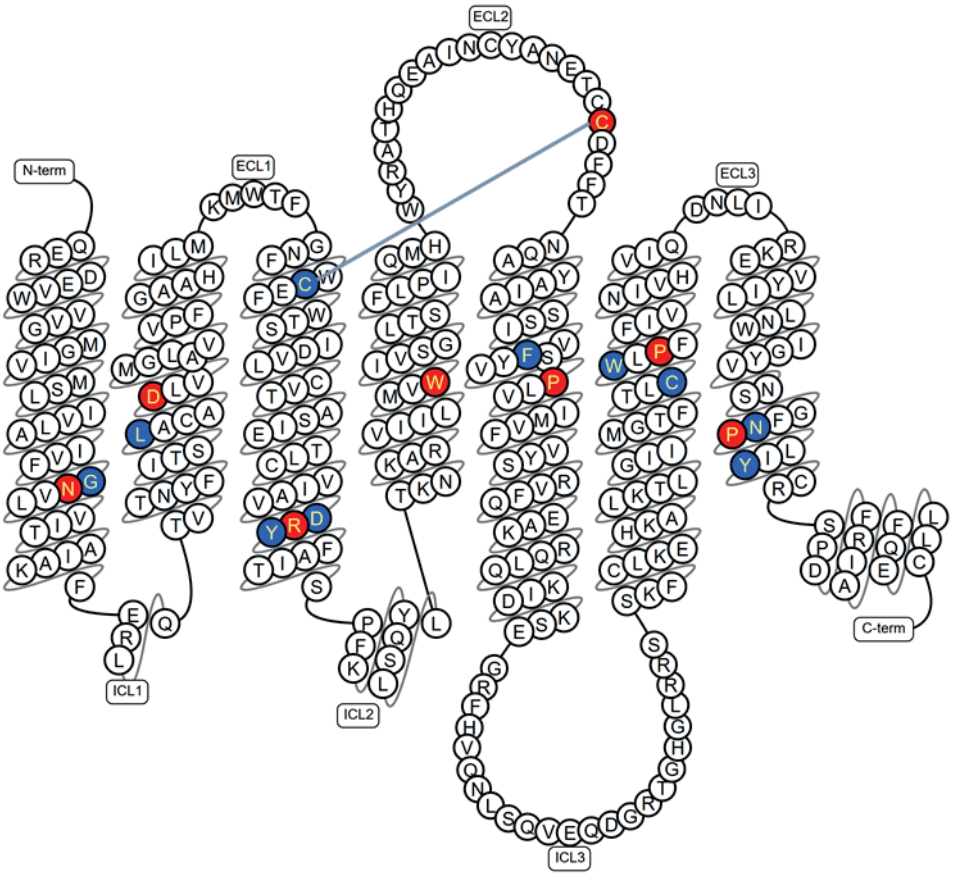

Figure 9. Conserved motifs (highlighted with red and blue), and pivot amino acids (highlighted with red) of TM domains and ECL2 of the rhodopsin family GPCRs, exemplified by the β_2-adrenoceptor. The cysteine bridge between C3.25 (TM3) and C45.50 (ECL2), common to many rhodopsin GPCRs, represented with a blue line. Termini not presented. Figure adapted from GPCRdb (Pándy-Szekeres et al., 2018).

Mostly, only transmembrane segments, less variable than the full sequence over evolution, are accounted for in GPCR evolution studies. In molecular phylogeny, gaps in sequence alignments have a meaning of an insertion-deletion event, in contrast to structural alignments, for example, where gaps indicate the lack of a 3D match (Carpentier and Chomilier, 2019). Tree reconstruction methods classically separate distance-based methods such as neighbour-joining or maximum likelihood and cladistics methods such as maximum parsimony, which aims towards ancestral sequences, and reconstruct trees following the parsimony principle, i.e., limit the number of evolutionary events.

Molecular phylogeny tends to be limited by noise that occurs over a long evolutionary time and the difference in divergence rates, which is typical when new receptors are created and need a new function. Upon duplication, the ancestral receptor often maintains the most conserved sequence throughout the species and tends to retain the original function (Ohno, 1970). Another limitation is that a limited number of taxa can be misleading in the general picture by omitting branching events.

Graphs-like representation – The basic local alignment search tool (BLAST) algorithm returns an expectancy value (e-value) when comparing two sequences, which depends on the length of the high-scoring pair-matching fragment and the number of matches in this region (Altschul et al., 1990). Pairwise matrices of e-values can then be combined into graphs, where the edges indicate close relationships. This type of study has been used, for example, to map GPCRs (Jekely, 2013; Bauknecht and Jékely, 2017).

Other sequence-based classification – In particular, Hidden Markov Models have been used for GPCR research sequence analysis to classify sequences for certain categories such as G protein coupling (Sreekumar et al., 2004). These techniques usually train a set of sequences using machine learning and apply the model.

Chromosomal mappings – Another technique used to reconstruct evolutionary histories is comparative genomics: genes are mapped to chromosomes, and the maps are used to follow possible duplications and deletions events. Many studies have used chromosomal mapping, particularly the complete mapping of neuropeptide receptors (Yun et al., 2015). More simply, only flanking genes can be used.

Studies on gene expression – Gene expression analyses can help map the tissues in which receptors are expressed and therefore provide insights into their functions. EST data searches can be useful to identify the expression of new receptors (Ritschard et al., 2019).

Pharmacology of ligand binding – Ligand preferences have been traditionally used to assess receptor subtypes, for example, adrenoceptor subtypes (Molinoff, 1984). When a new receptor is identified, it is primordial that it actually is activated by the endogenous ligand of the closest homologue. Nonetheless, this may be difficult for peptide receptors, where both the ligand and the receptor have evolved.

Coevolution of interacting proteins – A change in the ligand can be inferred, for example, from the absence or presence of a metabolizing enzyme. Such an enzyme can be catalysing amidation of peptides or the addition of a catalytic hydroxyl to adrenaline or dopamine (absence of the enzyme leads to tyramine and octopamine, found in insects) (Bauknecht and Jékely, 2017).

Structural studies on molecular evolution – There are instances where structural studies in 3D, often based on model receptors, have been used to study molecular evolution (Xhaard et al., 2006; Li et al., 2015).

2.6 Origin and diversity of GPCRs

GPCRs are classified as such based on their 7-TM-domain structure and coupling with heterotrimeric G-proteins. Several classification systems have been developed, and the best-known ones divide GPCRs into six (A–F) (Kolakowski, 1994) or five classes (GRAFS) (Fredriksson et al., 2003). Classes A–F comprise GPCRs throughout all species; class D and E receptors are absent in humans (Kolakowski, 1994). This classification was originally based on a fingerprinting method in which protein sequences were analysed by identifying conservative features from transmembrane helices (Attwood and Findlay, 1994). Recently, the more widely used classification system GRAFS divides GPCRs into five families: glutamate (class C), rhodopsin (class A), adhesion (class B), frizzled/taste2 (class F) and secretin (class B); GRAFS comprises only the human receptor classes (Fredriksson et al., 2003; Schiöth and Fredriksson, 2005) (Figure 10), so classes D and E are not included. The GRAFS classification system is based on phylogeny. Classes/families can be further divided into subfamilies and subtype groups based on phylogenetic relationships. The largest group, class A/rhodopsin, comprises about 700 receptors in humans, and about 400 of these are involved in olfaction. The phylogenetic tree of class A, rhodopsin-like receptors, forms four branches, α, β, γ and δ. Branches are further divided into 13 subfamilies: α-branch includes Amine, Melanocortin/EDG/cannabinoid/adenosine receptors (MECA), Prostaglandin,

Opsins and Melatonin receptors; β-branch includes Peptide receptors; γ-branch includes Chemokine, SOG and MCH receptors; and δ-branch includes leucine-rich repeat receptors (LGR), Mas-related (MRG), Purin and Olfactory receptors. Each of these subfamilies and branches shares a common ancestor receptor. Rhodopsin family GPCRs share conservative motifs in transmembrane parts, as well as family-specific motifs in extracellular and intracellular loops (Ruuskanen et al., 2004). The N-terminus is relatively short, and in most of the rhodopsin family receptors, the binding site is located in a pocket formed by α-helices. The glycoprotein hormone-binding receptors make an exception by binding with the N-terminus, similarly to other receptor classes (Moyle et al., 2004).

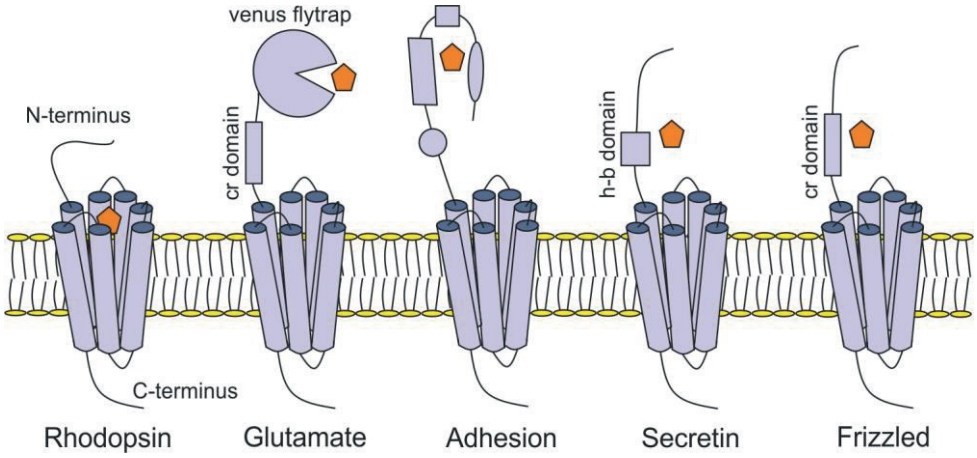

Figure 10. Structural differences in the ligand binding region between receptor families of the GRAFS classification system. The binding region of the rhodopsin family is inside the TM domain forming a pocket-like binding area, while the other families possess a binding region in their N-terminus and have remarkably longer N-termini compared to rhodopsin-like receptors. The orange pentagon represents ligands. CR domain: cysteine-rich domain; h-b domain: hormone-binding domain. Figure modified from (Krishnan, 2015).

The secretin family comprises 15 human receptors that participate in paracrine signalling and bind mostly large peptide ligands. Structurally, they share long cysteine-rich N-termini responsible for ligand binding. The adhesion class, with 33 human receptors, shares structural similarities with the secretin, and they have often been grouped together (class B) (Kolakowski, 1994). However, they most probably share distinct ancestors and should be grouped separately (Fredriksson et al., 2003; Bjarnadóttir et al., 2004). The adhesion class receptors also possess long N-termini,

often with epidermal growth factor domains (Mcknight and Gordon, 1996), whereas the N-termini of the secretin family receptors contain hormone-binding domains (reviewed in Lagerström and Schiöth, 2008).

Similar to the adhesion and secretin receptors, also the glutamate receptors possess a binding region in the N-terminus. The N-terminal binding area "Venus flytrap" is composed of conservative Ca^{2+}-recognising amino acids (Silve et al., 2005). The glutamate class is variable, and phylogenetic relationships have remained unclear (Fredriksson et al., 2003). The frizzled/taste 2 class consists of approximately 40 receptors in humans and is the latest addition to the GPCR protein family. The receptors share structural similarities, but their evolutionary lineages are probably different. The frizzled receptors are remarkably conserved through species, from invertebrates to vertebrates, while taste 2 receptors seem to be completely absent in invertebrates and, thus, most likely evolved later. Taste 2 receptors have more recently been classified as an independent class that evolved from the rhodopsin-like receptors (Nordström et al., 2008).

The glutamate family has been suggested as the oldest among GRAFS families (reviewed in Strotmann et al., 2011). Glutamate receptors are found in the slime mold *Dictyostelium discoideum*, whose origin has been tracked to about 600 million years ago. However, the glutamate-dependent signalling arose later. Ligand binding areas are thought to originate from the Venus flytrap and evolved to the N-terminal cysteine-rich domain and further to the binding cavity in the TM bundle. Cysteine composition is conserved between rhodopsin and glutamate families, indicating a close phylogenetic relationship. Rhodopsin family receptors are found in bilaterians, but also in cnidarians, indicating an ancient origin of these receptors. In addition to the glutamate receptors, the other ancient family is adhesion receptors, which also evolved before the metazoan evolution and are found in fungi and plants. The number of adhesion receptors varies remarkably between species, which indicates a fast gene duplication rate. Structurally, the TM region is surprisingly variable, probably due to N-terminal ligand binding. The secretin family is found in bilaterians but not in evolutionary older metazoa. Due to structural similarities to the adhesion family, it has been thought that secretin originated from the adhesion family. Frizzled receptors likely originated in multicellular organisms.

Following the same paradigm, sequence similarities within families are overall more conserved in ancient rhodopsin GPCR subfamilies like peptide (Larhammar and Salaneck, 2004) and amine receptors (Gloriam et al., 2005). More variability is

seen in more recent and fast-evolving families, like chemokine receptors (DeVries et al., 2006). The absence of these receptors in invertebrates outside the phylum *Chordata* also indicates a more recent evolutionary origin (DeVries et al., 2006; Kamesh et al., 2008).

GPCR subtypes are often referred to as ohnologues based on the theory of evolution by gene duplication (Ohno, 1970) (Figure 11). The main GPCR families arose around 1100–1400 million years ago, and their subtypes were divided 350–500 million years ago after whole-genome duplication in vertebrates. The vertebrate lineage went through two whole-genome duplications after the separation from the invertebrate lineage, leading to four subtypes of each ancestral GPCR (Ohno, 1970; Holland, 1999). Thus, as simplified, the repertoire of GPCRs in vertebrates is four times larger than in invertebrates. This two-round whole-genome duplication is often referred to as "2R", indicating two replications of the genome. It is hypothesised that 2R, which happened around the Cambrian explosion, is one of the main reasons for increased complexity among species and body functions (Ohno, 1970; Dehal and Boore, 2005). Ray-finned fishes went through additional whole-genome duplication (3R) (250–300 million years ago), leading to eight subtypes of each ancestral GPCR (Taylor et al., 2001). In principle, the scenario of the four/eight receptor subtype is not the reality; local duplications (fast-evolving genes) and deletions (pseudogenes) make the pattern of receptors much more complex. Only in rare cases, four or eight subtypes are actually seen.

It has been suggested that the larger repertoire of receptors has improved the ability for adaptation and has been favourable for biodiversity (Ravi and Venkatesh, 2008; Venkatesh et al., 2014). Not surprisingly, ray-finned fishes are the most rich vertebrates in the species: there are more than 27 000 living species, about half of all vertebrates (Nelson et al., 2016). However, there is no correlation between the gene number and the number of extinct species (Donoghue and Purnell, 2005), leading to a hypothesis of a highly "resilient" species with enhanced regulatory networks rather than the gene number directly affecting the species number (biodiversity of ray-finned fishes) after 3R (Levine and Tjian, 2003; Postlethwait et al., 2004; Crow and Wagner, 2006). Apart from 3R ray-finned fishes, there are 2R "basal teleost" species, e.g., spotted gar (*Lepisosteus oculatus*), which separated from the lineage of other ray-finned fishes before the third round of whole-genome duplication (Braasch et al., 2016; Pasquier et al., 2017). Thus, the genome duplications of spotted gar follow the same pattern as humans, following the theory of four GPCR subtypes per ancestral

receptor. The spotted gar genome has evolved relatively slowly and has remained notably conserved as compared with the other tetrapod genomes, bridging the differences between tetrapods and 3R ray-finned fishes. Sharks and other cartilaginous fishes also follow the four-subtype pattern (Venkatesh et al., 2014).

It is a generally accepted hypothesis that jawed vertebrates (*Gnathostomata*) have gone through the two/three rounds of whole-genome duplications. However, it is debated whether jawless vertebrates (*Agantha*) emerged before or after the second genome duplication (2R) (Dores, 2011; Caputo Barucchi et al., 2013; Venkatesh et al., 2014; Smith and Keinath, 2015; Sacerdot et al., 2018), but recent data indicate that the jawless vertebrates' lineage diverged from jawed vertebrates after 2R. Tunicates – which are invertebrates but part of the phylum *Chordata* – GPCRs arose prior to the whole-genome duplications but after the split of the chordates from the invertebrate lineage (Kamesh et al., 2008; Van de Peer et al., 2009), and thus, their GPCR repertoire is remarkably smaller than the repertoire of vertebrate GPCRs.

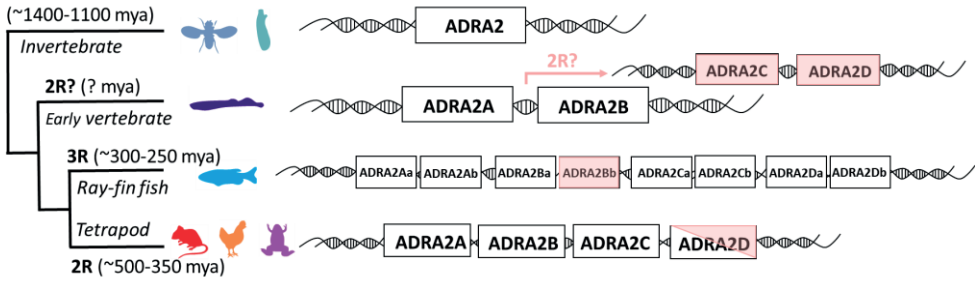

Figure 11. Schematic presentation of the evolution of α2-adrenoceptor subtypes (gene code ADRA2) through whole-genome duplications (1R, 2R, 3R). The figure describes the schematic pattern of the α2-adrenoceptor subtype composition in species groups and addresses the complexity of gene composition that rarely follows the pattern of four/eight ohnologues in vertebrate species. Red filling indicates local deletion and red arrow local duplication.

2.6.1 Identification of early neuropeptide signalling

The pairing of neuropeptide with invertebrates or early vertebrates has progressed from two angles: the discovery of orphan genomic sequences, either from the receptor (Sakurai et al., 1998) or from the peptides from databases, and the discovery of peptides from mass spectroscopy–based peptidomics (Shiraishi et al., 2019).

Based on the primary structure, the phylogeny of the neuropeptides is more difficult to predict than their cognate receptors (Grimmelikhuijzen and Hauser, 2012; Jekely, 2013; Mirabeau and Joly, 2013). Thus, the evolutionary analysis of neuropeptides from invertebrates and vertebrates has been challenging. The coevolution of neuropeptides and neuropeptide receptors can be used for predictions, but the pattern is not always straightforward. In some cases, distant receptors have acquired the same ligand binding preferences, and the origin cannot be easily concluded based on endogenous ligands. However, this is not the case with neuropeptides; it has been shown that neuropeptides and their cognate receptors follow similar phyletic patterns, which refers to long-range coevolution (Jekely, 2013; Mirabeau and Joly, 2013). Similarly to orexin receptors, high evolutionary pressure has also conserved prepro-orexin, as particularly orexin-A is highly similar among vertebrates (Figure 12). Orexin propeptide is found in most vertebrates, and the sequence has also been identified in many invertebrates, such as the cephalochordate *B. floridae* and the hemichordate *Saccoglossus kowalevskii* (Jekely, 2013; Mirabeau and Joly, 2013). Surprisingly, for *Petromyzon marinus*, a sea lamprey (*Chordata*, vertebrate), and for tunicates (*Chordata*, invertebrate), such as *C. intestinalis* and *Ciona savignyi*, orexin propeptide has not been identified, even though it is found at the "lower" and "higher" levels of taxa in the phylogenetic tree.

The problematic features of the neuropeptide precursor for sequence analysis are variability, repetitivity and shortness. Overall, the sequences of propeptides are often variable between species, actual neuropeptide sequences are too short for liable comparative sequence analysis, and similar motifs are often repeated even between distant neuropeptides. Thus, standard methods such as BLAST fail to identify evolutionary distant neuropeptides, so it is the case for invertebrate neuropeptides when considering a vertebrate query. Signal peptides are sometimes used as markers of neuropeptides (Semmens et al., 2016), but invertebrate neuropeptide precursors often include several, even dozens, of (copies of) neuropeptides, while vertebrate precursors range from one to a few (Wegener and Gorbashov, 2008). In addition, the pool of neuropeptides in different species groups is very heterogeneous.

Only in recent years, the gaps in the big picture of the evolution of neuropeptides have been bridged little by little (reviewed in Elphick et al., 2018). The species bridging the differences between deuterostomes and protostomes, such as echinoderms and tunicates, have a key role in broadening the knowledge of evolution and the ancient role of neuropeptides. Echinoderms and tunicates have features

typical of both deuterostomes and protostomes, which makes them unique and potential model organisms, e.g., in the neurosciences. Solving the origin and understanding the evolution of nervous systems might be key in better understanding complex nervous systems such as the human brain.

There is undisputed evidence that neuropeptides have emerged early, and various neuropeptides are also found in species with simple nervous systems, for example, non-bilaterian cnidarians (e.g., jellyfishes and sea anemones) (Grimmelikhuijzen et al., 1996). However, the link between cnidarians and vertebrate neuropeptides is more or less ambiguous. Neuropeptide precursors have also been reported in *Trichoplax adhaerens*, which lacks neurons (Jekely, 2013). The repertoire of neuropeptides or neuropeptide-like sequences are reported in many bilaterian invertebrate species from several subphylums, such as *Annelida* (Veenstra, 2011), *Mollusca* (Veenstra, 2010), *Arthropoda* (Wegener and Gorbashov, 2008), *Nematoda* (Nathoo et al., 2001; Li and Kim, 2010), *Echinodermata* (Semmens and Elphick, 2017), as well as *Tunicata* (Shiraishi et al., 2019, reviewed in Kawada et al., 2010). However, the repertoire of the reported cognate receptors is often larger than that of the reported neuropeptides, e.g., in *D. melanogaster* (Hewes and Taghert, 2001; Nathoo et al., 2001).

Parallel and comprehensive mapping of neuropeptides and neuropeptide receptors has provided insights into the endogenous ligands of orphan receptors, as well as clarified evolutionary origins of many receptor systems (Jekely, 2013; Mirabeau and Joly, 2013). Several neuropeptides have been suggested to be orthologous between the deuterostome and the protostome, indicating an ancient evolutionary origin of these neuropeptides. For instance, the association between the following deuterostome and protostome neuropeptides are described: neuropeptide S and CCAP, orexin and allatotropin, as well as galanin and allostatin A (deuterostome and protostome, respectively) (Mirabeau and Joly, 2013).

Allatotropin is an insect neuropeptide originally identified in *Manducta sexta*, where it stimulates the production and release of the juvenile hormones and modulates the circadian clock and myotropic activity (Kataoka et al., 1989). Understanding the connection between the insect allatotropin and the orexin systems led to the identification of the first invertebrate orexin. The precursor segment conserved between the protostome and non-Chordata deuterostomes is not found in vertebrate prepro-orexin (Jekely, 2013; Elphick et al., 2018). Additionally, orexin and allatotropin receptors group together in phylogenetic analysis (Yamanaka et al., 2008;

Jekely, 2013), and, e.g., the closest mammalian protein sequence of the *Manducta sexta* allatotropin receptor is the orexin receptor (Horodyski et al., 2011). Allatotropin is also suggested to have a role in feeding (Alzugaray et al., 2013).

Figure 12. Alignment of prepro-orexin peptide sequences of the vertebrates H. sapiens, M. musculus, G. gallus, X. laevis, Latimeria chalumnae (coelacanth) and D. rerio and the invertebrate B. floridae. The alignment has been manually modified to prevent gaps within secondary structure elements, as well as to align cysteines within the orexin-A. Cleavage sites of orexin-A and orexin-B are highlighted with grey; conserved amino acids are highlighted in light violet; the black boxes represents the orexin-A and orexin-B (or equivalent region). Stars indicate corresponding amino acids, colon conservative amino acid substitution, and dots indicate semi-conservative substitution and space non-conservative substitution.

3 Aims of the study

This thesis aims to close the gaps in knowledge by cartographing the repertoire of GPCRs in non-human vertebrates (Publication **I**) by studying a putative orexin receptor from the vase tunicate *C. intestinalis* (Publication **IV**) and by characterising the signalling pathways of small-molecular weight agonists in the human orexin receptor (Publication **II** and **III**).

The specific aims of the thesis were:

- ❖ to classify non-human vertebrate GPCRs (Publication **I**).
- ❖ to understand early branching and subtype conservation in the GPCR phylogenetic tree (Publication **I**).
- ❖ to characterise the *in vitro* features of the compounds Yan7874 and Nag26 and their effects on orexin receptor signalling pathways (Publications **II** and **III**).
- ❖ to demonstrate the relationship of the putative orexin receptor of *C. intestinalis* and the orexin system, as well as to identify its endogenous ligand(s) (Publication **IV**).

There are thus some unifying links among the publications:

- Publications **II-IV** focus on orexin receptors, their pharmacology, and the evolutionary aspects of the orexin system. They all aim to better understand agonism and characterise the relationships between compounds and pathways.
- Publications **I** and **IV** are a logical continuation one of each other, where **I** predicts a new receptor and **IV** experimentally tests it.

4 Materials and Methods

The methods used range from computational to molecular biology, pharmacology and phylogenetic and evolutionary assessments. They match the different levels considered: studies at the GPCR family level to research on selected receptors; and molecular biology in cell systems to atomistic modelling.

4.1 Computational methods

Computational methods were used in publications **I** and **IV**. The experimental section of publication **I** was completely computational, comprising sequence analysis and phylogenetic construction, while **IV** was both computational and experimental and involved homology modelling in addition to sequence analysis. The large-scale bioinformatics data collection and automated extraction of the TM regions in publication **I**, which required scripting, was conducted by the co-authors.

4.1.1 Sequence analysis

Generally, sequence analysis utilises the information of DNA, RNA or protein sequences and can be used to identify close homologies, prediction of phylogenetic relationships and evolution, and structures and functions.

4.1.1.1 Sequence retrieval and preparation

Sequence analysis in publication **I** was conducted on sequences from predicted transcripts and associated gene trees, which were automatically computed and provided by the Ensembl consortium. In Ensembl, the predicted transcripts are validated based on an automated detection of homologues, including searches in the EST database and the matching of curated human sequences. Automated alignments are constructed on the MUSCLE software (Zerbino et al., 2018). The automatically constructed gene trees are combinations of five outputs, i.e., maximum likelihood trees based on two different types of distances and neighbour-joining trees based on three types.

The Ensembl database, release 67 (Flicek et al., 2012), was first queried using the list of gene IDs of rhodopsin family non-olfactory GPCRs provided by the IUPHAR (Sharman et al., 2011). The gene information collected included a list of sequences, gene trees, associated sequence alignments, and associated data. In most cases, closely-related GPCRs have been grouped by Ensembl inside a single gene tree, and in later releases, e.g., release 92 (Zerbino et al., 2018), in single subtrees. As a result, 71 groups containing unique receptors from 53 vertebrate genomes, from mammals to the lamprey, 2 tunicate genomes, 2 invertebrate genomes, and about 14 000 amino acid sequences were retrieved.

Retrieved data were curated, even though Ensembl only rarely includes alignment errors since rhodopsin GPCR sequences are all closely related. However, individual translational errors may occur, or some amino acid sequences are fragmental or have a mispredicted N- or C-terminus. Additionally, multiple occurrences of a given gene – seemingly duplicates – may be found. All the TM regions of the sequence alignments of 71 retrieved groups were visualised by the MEGA5 package (Tamura et al., 2011) to ensure that sequences were properly aligned within each group. Additionally, pivot positions of all seven TMs in the alignments (see pivot amino acids in 2.5 Methods to study molecular evolution) were annotated in the sequence alignments, and each (putative) subtype group inside the 71 gene trees was manually annotated.

One human sequence for each of the 71 gene trees was used as a reference (often well-characterised, appropriately named, and preferably not including missing regions in their amino acid sequences), and the quality of the rest of the sequences was controlled based on the reference. The quality requirements were a minimum of 10 %ID and a maximum of four deletions (gaps) in each of the seven TMs. These criteria allowed the elimination of the mistranslated sequences and fragments, as well as the inclusion of the more distant invertebrate sequences and misaligned ones, if any. More stringent criteria (<15%) led to overall similar results (Rinne, unpublished data). After the curation, the pool comprised 11 000 sequences. Subsequently, the 71 groups of closely-related receptors were assembled into 31 clusters based on their earlier classification of GPCRs (Fredriksson et al., 2003), comprising all rhodopsin main families, Amine (3 clusters), Opsins (1), Prostaglandin (1), Melatonin (1) and MECA (1), Peptide (1 cluster), SOG and MCH (combined into 1 cluster) and Chemokine (1), MAS (1), Glycoprotein (1) and Purin (1). Some receptors not identified in the earlier classification were assigned to families based on the Ensembl

automated trees. The remaining clusters of receptors, smaller groups with no associated families and mostly orphan receptors, were treated independently.

Data were also used to examine the sequence conservation between subtype groups and for mapping invertebrate and early vertebrate sequences. Subtype conservation was determined by calculating the %ID of 4 vertebrate species in relation to the human reference sequences: mouse (*Mus musculus*), chicken (*Gallus gallus*), frog (*Xenopus tropicalis*) and zebrafish (*D. rerio*). Additionally, invertebrate sequences were mapped to receptor groups based on their %ID in relation to human sequences.

The BLAST of Ensembl (Zerbino et al., 2018) was used to identify the putative *C. intestinalis* prepro-orexin (CiPPO). BLAST is developed to compare a query sequence to a library of sequences and to identify structurally related sequences. This has proven to be very challenging, perhaps due to lack of the correct gene in databases. For the successful run, the BLAST search was conducted in Ensembl-predicted transcripts with peptide sequences of several known prepro-orexin segments containing the signal peptide and orexin-A with the cleavage site against family *Ciona* nucleotide sequences; thus, TBLASTN was used. The search was adjusted for distant homologies and restricted with penalties for opening (10) and extension (3) in the Blosum45 matrix, and low complexity regions were filtered. The first hit, in *C. savignyi*, resulted from the *D. rerio* prepro-orexin segment, while no hits were found in *C. intestinalis*. The first hit was further queried against *C. intestinalis* and adjusted to close homologies, resulting in a hit, the putative CiPPO in *C. intestinalis*.

4.1.1.2 Phylogenetic reconstruction

In publication **I**, phylogenetic trees were constructed for each of the clusters of the curated sequences using the package MEGA5 (Tamura et al., 2011). Sequences were first realigned within each of the 31 clusters using ClustalW, and the alignments visually examined the TM regions to eliminate gross misalignments. Phylogenetic trees were then constructed using the neighbour-joining method combined with a 1000-fold bootstrap analysis to assess the robustness of the tree construction. These trees were compared with the Ensembl trees available in May 2015 (Cunningham et al., 2014) and October 2018 (Zerbino et al., 2018). The trees were found to be in good

agreement, even though the available number of genomes had almost doubled by 2018.

4.1.2 Homology modelling

Homology modelling refers to the prediction of molecular 3-dimensional structures of a homologous protein based on the template protein with a known 3-dimensional structure.

For publication **IV**, homology modelling was conducted to obtain the 3-dimensional structure of the putative *C. intestinalis* orexin receptor (CiOX) based on the structure of human OX_2 (PDB: 4S0V) (Yin et al., 2015) and to identify structural similarities and differences. The amino acid sequence of the putative CiOX was retrieved from Uniprot and compared with both human receptor subtypes. It shows a slightly higher %ID compared to OX_2, thus OX_2 was selected as a template. The sequence of CiOX was manually aligned with OX_2, and the absent structures, most of the ECL and ICL domains of OX_2, were also deleted from the CiOX sequence. The structure prediction was conducted with MODELLER version 9.14 (Webb and Sali, 2014) and visualised by PyMOL (PyMOL 1.7.0.0 Schrödinger).

A second homology model was constructed to examine the structure of the putative Ci-orexin-A. Based on the similarities in the primary structure of Ci-orexin-A and human orexin-A, a homology model was constructed using human orexin-A (PDB: 1R02) (Kim et al., 2004) as a template. One hundred models were constructed and visualised to ensure a similar structural folding to human orexin-A. The most promising structure was selected for publication **IV**.

4.2 Pharmacological assays

Pharmacological assays were used in publications **II**, **III** and **IV**. The (putative) orexin receptor agonists were the following: orexin-A, Nag26 (publication **II**), orexin-A, Yan7874 (publication **III**) and orexin-A, orexin-B, [Ala^{11},D-Leu^{15}]-orexin-B, the 19-amino acid-long C-terminal fragment of human orexin-A (orexin-A_{15-33}), Nag26 (synthesised in publication **II**, later commercially available), Yan7874 (commercially available) and the putative orexin peptide of *C. intestinalis* (Ci-orexin-A) (publication **IV**). Four different variants of the putative orexin peptide of *C.*

intestinalis (Ci-orexin-A) were tested: a short variant (18 amino acids long C-terminal fragment) and a long variant (43 amino acids, full peptide), amidated at the C-terminus, both either acetylated or not acetylated at the N-terminus. Orexin-A, orexin-B, [Ala^{11},D-Leu^{15}]-orexin-B and orexin-A$_{15-33}$ indicate human peptides unless otherwise noted.

Inhibitors were used to verify pathway-specific activation, e.g., orexin receptor antagonist TCS-1102 and almorexant, Gq-specific inhibitor UBO-QIC and MAPK/ERK kinase 1 inhibitor U0126 (Table 4).

*Table 4. Summary of the pharmacological tools used in publications **II**, **III** and **IV**.*

Tool	Use	IUPAC name
CTx	G_S activator	
Fluorescein diacetate	Stains viable cells	(6'-acetyloxy-3-oxospiro[2-benzofuran-1,9'-xanthene]-3'-yl) acetate
GF109203X	PKC inhibitor	2-(1-[3-dimethylaminopropyl]-1*H*-indol-3-yl)-3-(1*H*-indol-3-yl)-maleimide
Hoechst 33342	Stains all nuclei	2-(4-ethoxyphenyl)-6-[6-(4-methylpiperazin-1-yl)-1H-benzimidazol-2-yl]-1H-benzimidazole
IBMX (3-isobutyl-1-methylxanthine)	Phosphodiesterase inhibitor	1-methyl-3-(2-methylpropyl)-7H-purine-2,6-dione
PTx	G_i inhibitor	
Propidium iodide	Stains necrotic cells	3-(3,8-diamino-6-phenylphenanthridin-5-ium-5-yl)propyl-diethyl-methylazanium;diiodide
PrestoBlue/Reazurin	Stains viable cells	7-hydroxy-10-oxidophenoxazin-10-ium-3-one
TAMRA	Fluorescent label	2-[3-(dimethylamino)-6-dimethylazaniumylidenexanthen-9-yl]-5-(2,5-dioxopyrrol-1-yl)benzoate
TCS-1102	Potent DORA	N-biphenyl-2-yl-1-{[(1-methyl-1H-benzimidazol-2-yl)sulfanyl]acetyl}-L-prolinamide
U0126	MAPK/ERK kinase 1 inhibitor	(2Z,3Z)-bis({amino[(2-aminophenyl)sulfanyl]methylidene}) butanedinitrile
UBO-QIC	G_q inhibitor	L-threonine,(3 R)-N-acetyl-3-hydroxy-L-leucyl-(aR)-a-hydroxybenzenepropanoyl-2,3-idehydro-N-methylalanyl-L-alanyl-N-methyl-L-alanyl-(3 R)-3-[[(2 S,3 R)-3-hydroxy-4 methyl-1-oxo-2-[(1-oxopropyl)amino]pentyl]oxy]-L-leucyl-N,O-dimethyl-,(7→1)-lactone (9CI)
YO-PRO-1 iodide	Stains membrane damaged cells	4-[(3-methyl-1,3-benzoxazol-2(3 H)-ylidene)methyl]-1-[3-(trimethylammonio)propyl] quinolinium diiodide

4.2.1 Cell culture and medium

In publications **II**, **III** and **IV**, recombinant cell systems were used. In publications **II** and **III**, small-molecule orexin receptor agonists were pharmacologically characterised in three different CHO-K1 cell lines: CHOs expressing human OX_1 (CHO-OX_1) and OX_2 (CHO-OX_2) receptors as well as wild-type CHO cells (CHO-ctrl). CHO cells were cultured in Ham's F12 nutrient mixture (GibcoGibco/Life Technologies, Paisley, UK) supplemented with 10% fetal calf serum, 10 nM HEPES, 100 U/mL penicillin G (Sigma Chemical Co., St Louis, MO, USA) and 80 U/mL streptomycin (Sigma) on plastic cell culture dishes (Greiner Bio-One GmbH, Frickenhausen, Germany).

In publication **IV**, a recombinant HEK-293 cell line was used. Particularly, the Flp-In™ T-REx™ 293 (Thermo Fisher Scientific, Waltham, MA, USA) host cell line was used for the generation of stable cell lines expressing human OX_1 fused with green fluorescent protein (OX_1-GFP), human OX_2-GFP and *C. intestinalis* putative orexin receptor CiOX-GFP. These HEK cells were cultured in high-glucose Dulbecco's Modified Eagle Medium, high glucose (Biowest, Nuaillé, France) supplemented with 10% fetal calf serum, 1% Glutamax (Gibco), 10 mM HEPES, 100 U/mL penicillin G, 80 U/mL streptomycin, Blasticidin (15 µg/mL) and, for the maintenance of the host cell line Zeocin (100 µg/mL) and for transfected cell lines, hygromycin B (100 µg/mL).

HEPES-buffered medium (HBM) containing 137 mM NaCl, 5 mM KCl, 1.2 mM $MgCl_2$, 0.44 mM KH_2PO_4, 4.2 mM NaHCO3, 1 mM $CaCl_2$, 10 mM glucose, 20 mM HEPES, and adjusted to pH 7.4 with NaOH, was used as an experimental buffer. Assay buffers were further completed with assay-specific supplements (see chapters 4.2.2, 4.2.3, 4.2.4, 4.2.5, 4.2.6, Table 4).

4.2.2 Ca^{2+} elevation

Intracellular calcium elevation is widely used to study the activation of G_q-coupled GPCRs (and many others). Due to a rapidly seen response, fluorescence-based Ca^{2+} mobilisation assays are a beneficial tool for screening both agonists and antagonists. In pre-incubation with a fluorescent calcium probe, probenecid may be added to loading dye to maintain an intracellular calcium probe by inhibiting anion transporters.

For Ca^{2+} elevation studies, the FlexStation 3 fluorescence plate reader (Molecular Devices, Sunnyvale, CA, USA) was used. Cells were plated on black, clear-bottom 96-well plates, they were pre-coated with polyethyleneimine (PEI; 25 µg/ml, 1 h, 37 °C), and assays were run on adherent cells. On multiwell plates, the growth medium was replaced with the loading solution, the FLIPR Calcium 4 (in publications **II** and **III**) or 5 (in publication **IV**) Assay Kit dissolved in and diluted with HBM and supplemented with 1 mM probenecid (and in publication **IV** 0.2% bovine serum albumin). Plates were incubated for 60 min at 37 °C. Inhibitors (in publications **II**–**III**: TCS-1102, UBO-QIC; in publication **IV**: TCS-1102, UBO-QIC, almorexant, Ci-orexin-A) were added 30 min prior to stimulation. Intracellular Ca^{2+} levels (fluorescence changes) were measured for 150 s (stimulation at 30 s) with excitation at 485 nm and emission at 525 nm (recording every 1.3 s) at 37 °C.

4.2.3 PLC activity

PLC activation was assessed by measuring cytosolic inositol phosphates, and in particular, the specific indicator of PLC activity is IP_3. Cells are pre-incubated with [^{3}H]-inositol in order to label the membrane phosphoinositides, which are hydrolysed to inositol phosphates upon PLC activation. The experimental buffer is supplemented with LiCl to inhibit cytosolic inositol monophosphatase.

In publications **II** and **III**, PLC activity was conducted with adherent CHO cells plated on clear 48-well or 96-well plates that were pre-coated with PEI as for the Ca^{2+} assays. Cells were allowed to grow for 24 h, 20 h of which were in the presence of the [^{3}H]-inositol (3 µCi/mL) label. The growth medium was removed, and cells were incubated in HBM supplemented with 10 mM LiCl for 30 min at 37 °C. Also, the inhibitor DORA TCS-1102 and the G_q inhibitor UBO-QIC were added at this stage. After this incubation, stimulants were added. Reactions were stopped after 10- or 30-min stimulation by rapidly replacing the medium with ice-cold perchloric acid (0.4 M) followed by freezing, careful re-thawing and neutralisation with 0.36 M KOH + 0.3 M $KHCO_3$. Insoluble fragments were spun down, and the supernatants were collected for anion exchange chromatography to isolate the total inositol phosphate fraction. Radioactivity was measured by scintillation counting (HiSafe 3 scintillation cocktail and Wallac 1415 liquid scintillation counter; PerkinElmer).

4.2.4 AC activity

AC activity was used to assess the activation of the alternative signal pathways of the human orexin receptors G_S and G_i. Since G_S activates and G_i inhibits AC, the effect of one needs to be prevented when studying the other. Cholera toxin (CTx) locks G_S in its GTP-bound state, which leads to G_S stimulation and the saturation of AC. This enables the measurement of the inhibitory effect of the activated G_i on AC. In turn, pertussis toxin (PTx) locks G_i in its GDP-bound state and thus prevents the inhibitory effect of G_i, which enables the measurement of the activating effect of the activated G_S on AC. For the determination of AC activity, cells were pre-incubated with [^{3}H]-adenosine in order to label cellular ATP. HBM was supplemented with a cyclic nucleotide phosphodiesterase inhibitor (3-isobutyl-1-methylxanthine, IBMX; to prevent hydrolysis of cAMP) and a protein kinase C inhibitor (GF109203X) because PKC can activate AC independently from orexin receptors in CHO cells (Holmqvist et al., 2005; Kukkonen, 2016). Upon stimulation, the AC activator forskolin was added to enhance the production of cAMP.

In publications **II** and **III**, AC activity was measured. Adherent CHO cells were plated on clear 48- or 96-well plates (PTx and CTx treatments on different plates). After 8 hours, the cells were treated with PTx and after 24 hours with CTx, and after 48 hours from plating, the cells were labelled with [^{3}H]-adenosine (5 μCi/mL) and incubated for 2 hours. After the incubation, the cells were washed once with phosphate-buffered saline (PBS) and HBM supplemented with IBMX (500 μM), and GF109203X (3 μM) was added to the cells and incubated for 30 min at 37 °C. Forskolin ± stimulants were added followed by an additional 10-min incubation. Reactions were stopped by rapidly replacing the medium with ice-cold perchloric acid (0.33 M) followed by freezing. Samples were thawed and insoluble fragments were spun down. [^{3}H]ATP + [^{3}H]ADP and [^{3}H]cAMP fractions were isolated from the supernatant by sequential Dowex–alumina chromatography. Radioactivity was determined with scintillation counting as in PLC activation. The conversion of [^{3}H]ATP to [^{3}H]cAMP was calculated as the percentage of the total eluted [^{3}H]ATP + [^{3}H]ADP.

4.2.5 Elk-1 reporter gene assay

A gene promoter activity reporter assay, a luciferase activation of ETS Like-1 protein (Elk-1) was determined to assess G_q-dependent activation of orexin receptors in publication **II**. The transcription factor Elk-1 is activated by ERK1/2 but also by other MAPKs (Price et al., 1996). Involvement of ERK1/2 was verified with U0126, an inhibitor of MAPK/ERK kinase 1 (an upstream activator of ERK1/2) (Jäntti et al., 2013).

Transient transfection was used for the co-transfection of plasmids pSG-GalElk-1 (fusion of the dimerisation domain of the Elk-1 and the DNA-binding domain of the yeast transcription factor Gal4) (Kortenjann et al., 1994), pGL3 G5 E4 Δ38 (5 × Gal4 binding site controlling firefly luciferase expression) (Kamano et al., 1995), pRL-TK (Renilla luciferase under Herpes simplex virus thymidine kinase promoter) (Promega, Madison, WI, USA) and empty pUC18 (for the optimisation of the DNA amount) in CHO-OX_1 and CHO-OX_2 cells plated on PEI-coated clear 96-well plates at 40% confluence. Transfection was conducted in Ham's F-12 with 0.312 mg/cm^2 DNA and 0.74 ml/cm^2 FugeneHD (Roche, Mannheim, Germany), with optimised conditions [0.6% (w/v) pSG-GalElk-1, 50% pGL3 G5 E4 Δ38, 10.8% pRL-TK, and 38.6% of pUC18]. Five hours after the transfection, the transfection media were removed and the cells were washed and serum-starved overnight in serum-free media.

Next day cells were stimulated in fresh serum-free media. Also, inhibitors (TCS-1102, UBO-QIC and U0126) were added at this stage. After 5 hours, the cells were lysed and the samples frozen. Luminescence activity was measured with the Dual-Luciferase Reporter Assay system (Promega) according to the manufacturer's protocol, with GloMax 20/20 luminometer (Promega). Firefly luciferase, an Elk-1-specific signal, was normalised to an average Renilla luciferase signal (total luciferase activity).

4.2.6 Cell viability

In publications **II** and **III**, cell viability was studied to assess the orexin receptor–specific and non-specific effects of compounds Nag26 and Yan7874. Fluorescent dyes were used to stain the cells (Table 4). Hoechst 33342 stains all nuclei and can be used to indicate the total cell number and to assess viable/apoptotic

cells based on the morphology of nuclei. YO-PRO-1 iodide has been claimed to stain apoptotic cells, but it seems to stain already membrane-damaged cells (own unpublished data), i.e., cells in the primary and secondary necrosis stages. Propidium iodide (PI) is also used to dye necrotic cells. PrestoBlue® is used to stain viable cells by reducing non-fluorescent resazurin to red fluorescent resorufin in the mitochondria. Another dye for viable cells, fluorescein diacetate (FDA), is also non-fluorescent but is hydrolysed in the cytoplasm of viable cells to fluorescent fluorescein.

In publication **II**, the cells were plated on black, clear-bottom 96-well plates and cultured for 24 hours. Then, cells were washed and incubated for 10 min in PBS and stimulated in a serum-free medium for 72 hours. After stimulation, cells were stained with Hoechst 33342 (10 μM), YO-PRO-1 iodide (0.1 μM) and PrestoBlue (0.1 μM) for 25 min at 37 °C. FlexStation 3 was used for reading the fluorescence at wavelengths 352 nm/455 nm (Hoechst 33342), 480 nm/515 nm (YO-PRO-1 iodide) and 550/590 nm (PrestoBlue®).

In publication **III**, the cells were plated on black clear-bottom 96-well plates and treated the following day. Staining was conducted 24 hours later with Hoechst 33342 (10 μM), PI (1 μM) and FDA (1 μM) for 20–30 min at 37 °C. Cell viability was assessed with both quantitative and qualitative methods. In the quantitative method, FlexStation 3 was used to determine the total cell number, viable cells and necrotic cells at wavelengths 352 nm/455 nm (Hoechst), 480 nm/525 nm (FDA) and 538/617 nm (PI). The cells were also assessed with fluorescent microscopy (Nikon TE2000 fluorescence microscope with 20×/0.75 air objective and the images acquired by an Andor iXon 885 electron-multiplying charge-coupled device camera (Andor Technology Ltd., Belfast, UK) under the control of Nikon NIS Elements AR software). Additionally, morphological features of the unstained cells were assessed by phase-contrast microscopy (Olympus CKX41 microscope with attached Canon EOS 600D digital camera) once every 24 h up to 72 h.

4.2.7 Membrane localisation of labelled orexin-A

For publication **IV**, binding of the fluorescent TAMRA (5- and 6-carboxytetramethylrhodamin)-labelled orexin-A (TAMRA-orexin-A) to the plasma membranes of HEK-OX_1-GFP and HEK-CiOX-GFP cells and the host cells Flp-In T-REx 293 (negative control) was assessed with fluorescent microscopy (Nikon

TE2000 fluorescence microscope with 60×/1.49 oil objective and the images acquired with an Andor iXon 885 electron-multiplying charge-coupled device camera under the control of Nikon NIS Elements AR software).

The cells were plated on PEI-coated 25 mm circular glass coverslips (Menzel-Gläser, Braunschweig, Germany), cultured overnight and treated with doxycycline to induce receptor expression (see 4.3.2 Transfection and expression). After 24 hours from the doxycycline addition, the cells were washed once with HBM, which was also used as the experimental buffer. For the imaging, cells were exposed first at 540/25 nm excitation, and emitted light was collected through a 565 nm dichroic mirror and a 605/55 nm band-pass filter to observe TAMRA. Right after, cells were exposed to 480/30 nm excitation, and emission was collected through a 510 nm dichroic mirror and a 535/40 nm band pass filter to observe GFP.

During the imaging, TAMRA-orexin-A was added to cells directly to the chamber, and its binding to the cells was followed consecutively every minute or every five minutes, depending on the aim of the experiment. Incubation time with TAMRA-orexin-A ranged from 15 min to 45 min. Additionally, some cells were pre-incubated in the chambers for 30 min with inhibitors (TCS-1102 and almorexant) or the putative Ci-orexin-A-peptide to assess their effect on TAMRA-orexin-A binding. Ci-orexin-A-peptide was also added after TAMRA-orexin-A incubation in order to assess the possible displacement. All incubations were carried out at room temperature.

4.3 Molecular biology

Molecular biological methods were only used in publication **IV**.

4.3.1 Cloning

The received cDNA (https://dna.brc.riken.jp/en/cloneseten/ciona_est_en) of CiOX was inserted into the pBluescriptII SK(-) vector and further transferred to the pCEP vector, which enabled C-terminal fusion to the fluorescent protein EGFP.

The cDNA of CiOX fused with C-terminal GFP was transferred into the Flp-In system expression vector pcDNA5/FRT using restriction endonuclease cloning. Additionally, wild-type human OX_1-GFP and OX_2-GFP (constructed in Putula et al.,

2011b) were sub-cloned into the same expression vector. The cloning was planned using the software SerialCloner2.6 (Perez, 2004).

4.3.2 Generating stable cell lines

Constructed plasmids, pcDNA5/FRT-OX_1-GFP, -OX_2-GFP and -CiOX-GFP, were used to generate stable cell lines in Flp-In™ T-REx™ 293 cells. This system ensures homogenous expression since the transcriptionally active locus of each genome contains just a single integrated flippase recognition target (FRT; site-directed recombination) site. Co-transfection of the gene of interest in the Flp-In expression vector with the Flp recombinase vector pOG44 leads to targeted integration of the gene of interest into the FRT site, thus in the same locus in each cell. Correct gene integration results in the loss of Zeocin resistance and the gain of hygromycin resistance.

For the transfection, host cells, Flp-In T-REx 293, were seeded in 6-cm plastic culture dishes. Cells were cultured to 60–80% confluence and washed with PBS, and the medium was replaced with fresh antibiotic-free growth medium. Cells were transfected with the gene of interest and pOG44 in the ratio 1:10, and 0.134 µg/cm^2 DNA in total. The transfection reagent GeneJuice® (Merck, Darmstadt, Germany) was used according to the manufacturer's protocol, with the µL transfection reagent / µg DNA ratio 3:1. The next day, the transfection medium was removed, cells were washed with PBS and fresh growth medium without Zeocin or hygromycin B was added. On the next day, cells were detached, reseeded in 25% confluence and incubated for 2–3 hours to ensure the attachment of the cells. The medium was replaced with the selection medium containing hygromycin B (100 µg/mL). The selection medium was changed to fresh every three or four days until visible cell foci were formed. For each cell type, four cell foci were picked, expanded and screened for doxycycline-induced gene expression and Zeocin sensitivity. The best foci of each cell line were selected for continuation.

The expression of GFP (fused to the orexin receptors) was assessed for several different doxycycline concentrations and added to cells 24 hours after the plating. The expression was qualitatively assessed by fluorescent microscopy (described in 4.2.6 Cell viability) at 24 hours and 48 hours. Expression levels of CiOX were equally high at 100 ng/mL and 1 µg/mL doxycycline; thus, the lower concentration was selected for the following experiments.

4.4 Data analysis

All experiments were conducted at least in triplicate and from several batches of cells. Data are presented as mean ± SEM. Statistical comparisons were conducted with Student's non-paired or paired two-tailed t-test with Bonferroni correction for multiple comparisons in publications **II** and **III**, and with oneway-ANOVA followed by Bonferroni correction for multiple comparisons in publication **IV**. Significances were as follows: ns (not significant; $P > 0.05$), * $P < 0.05$, * * $P < 0.01$, * ** $P < 0.001$.

Microsoft Excel and Prism 5 (GraphPad Software Inc.) were used for data visualisations and analyses. Statistical analyses and non-linear curve-fitting were conducted with Prism 5 (GraphPad Software Inc., La Jolla, CA, USA). Curve-fitting equations were as follows:

$$\text{response} = \text{basal} + \frac{[\text{ligand}]^{n_h} \times \text{response}_{max}}{[\text{ligand}]^{n_h} + {EC_{50}}^{n_h}} \qquad \text{(Eq. 1)}$$

$$\text{response} = \text{basal} + \frac{[\text{ligand}]^{n_{h1}} \times \text{response}_{max}}{[\text{ligand}]^{n_{h1}} \left(1 + \frac{[\text{ligand}]^{n_{h2}}}{{K_i}'^{n_{h2}}}\right) + {EC_{50}}^{n_{h1}}} \qquad \text{(Eq. 2)}$$

$$Y = \text{basal} + \frac{(\text{response}_{max} - \text{basal})}{1 + 10^{(\text{Log}EC_{50} - X)}} \qquad \text{(Eq. 3)}$$

Eq.1 was used for curve-fitting of all concentration response data in publications **II** and **III**, Eq. 2 was used for alternative curve-fitting of the Elk-1 data (uncompetitive inhibition) in publication **II**. In publication **III**, Ca^{2+} concentration response data were fitted according to Eq. 3.

5 Results and Discussion

5.1 Classification of non-human GPCRs (publication I)

The challenge of the first publication was constantly updating the Ensembl database. In the timespan of publication **I**, Ensembl got updated 25 times. Additionally, the growing sequence of data reorganised the structure of phylogenetic trees of some receptors, but at the same time, it increased the liability of the phylogenetic relationships. Especially the genome additions from basal teleosts (2R), such as the spotted gar, were pivotal in the final classification of putative receptor subtype clades.

Based on the retrieved sequence data and later updates (Ensembl releases 67 (Flicek et al., 2012) and 92 (Zerbino et al., 2018)), 142 receptor clusters with no human orthologue in rhodopsin-like GPCRs were identified. To the best of our knowledge, 69 of these are not found in the literature. Of these, 114 clusters were mapped to already known and named subtype groups, while 28 were suggested to be orphan receptors and thus not part of any known subtype group. Within the main families, we identified a total of 14 Amine receptor groups with no human orthologue (5 novel / 1 orphan), MECA 6 (4 / 4), Opsins 19 (1 / 0), Prostaglandin 2 (1 / 0), Melatonin 4 (0 / 0), Peptide 23 (8 / 1), Chemokine 15 (8 / 4), SOG & MCH 14 (5 / 2), LGR 2 (0 / 0), MRG 3 (0 / 0), Purin 33 (30 / 12) and Others 7 (7 / 4). Phylogenetic reconstruction was used for constructing the descriptive template trees with newly identified and annotated receptors.

Ray-finned fishes (3R) are represented in most of the identified receptor groups (134) (Table 5). When moving towards humans, fewer novel receptors are presented by the groups of species, amphibians possess only 69 of the identified receptors and placental mammals only 11 of the novel receptors. When considering the genome duplication theory and the common research interests, the findings are not too surprising. Most of the new receptors were found in two or more groups of species, reflecting relatively strong data support. Putative receptors represented by only one sequence from one species were not considered. Not surprisingly, most of the receptors were found from fast-evolving recent main families, such as Purin and Chemokine.

Table 5. Ensembl release 91 included 90 vertebrate genomes and 4 invertebrate genomes. Species group–specific counts of novel putative GPCRs with no-human orthologues. Table adapted from (Rinne et al., 2019).

	Lamprey (a)	Fishes (*Actinopterygii*)		Fishes (*Sarcopterygii*)	Amphibian	Reptiles	Birds	Mammals		
Representative species	*Petromyzon marinus*	Cypriniformes (*Danio rerio*); Characiformes (*Astyanax mexicanus*)	*Lepisosteus oculatus* + *Takifugu rubripes* and 7 others	*Latimeria chalumnae*	*Xenopus tropicalis*	*Anolis carolinensis*; *Pelodiscus sinensis*	*Gallus gallus* + 4 others	Monotreme	Marsupial	Placental
Whole genome duplication	1R/2R?	3R	3R, except *Lepisosteus oculatus* 2R	2R	2R	2R	2R	2R	2R	2R
Pictogram										
Number of genomes	1	2	9	1	1	2	5	1	3	65
Paralogous GPCR clades identified	22	118	134	95	69	74	52	16	19	11

Sequence data were also used to assess conservation through species. The human reference sequence of each receptor group was compared to the orthologous gene of a mouse (*M. musculus*), a chicken (*G. gallus*), a frog (*X. tropicalis*) and a zebrafish (*D. rerio*). The common hypothesis suggests the high conservation of the "original" receptor throughout the species (before the whole-genome duplications) within the subtype group (Ohno, 1970). The results show an apparent variety between subtype conservation through species, and in many subtype groups, the most conserved subtype is clearly identified. Additionally, subtype conservation varies among the main branches; α- and β- branches are generally more conserved than fast-evolving γ- and δ-branches.

While the main aim of the study was to identify novel non-human GPCRs and map them to known subtype groups, the study also describes the mapping of human orphan receptors to already known subtype groups. Nineteen orphan receptors seem to be stably constructed in Ensembl, e.g., in the peptide family, GPR39 clusters together with neurotensin receptors (NTSR), and GPR37 and -37L1 cluster together with endothelin receptors. This mapping might give new insights into the function and ligand preferences of the human orphan receptors. Additionally, invertebrate (*D. melanogaster, C. elegans, C. intestinalis, C. savignyi*) and early vertebrate (*P. marinus*) GPCR genes were mapped to known subtype groups based on the %ID of the TM regions of human orthologues (Rinne, unpublished data; Figure 13). Thus, this data can be utilised in identifying the evolutionary origin of many receptors; e.g., the putative orexin receptor of *C. intestinalis* shows the highest %ID in human OX_2 (44%) and the second highest to human OX_1 (43%).

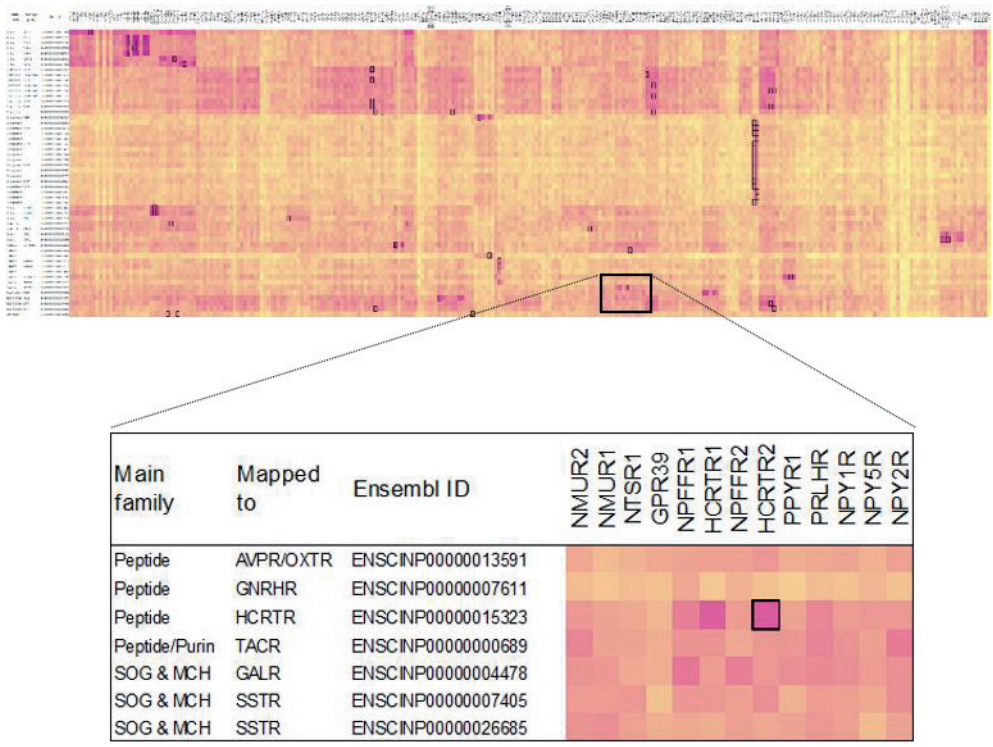

Figure 13. Above, the heat map of %ID of C. intestinalis GPCRs in relation to the human sequence of each receptor clade. Below, the zoom-in from above, C. intestinalis receptors mapped to human peptide receptors, and the hit of the mapped ENSCINP00000015323 to the human OX_2 orexin receptor (HCRTR2). The black squares represent the highest %ID of each row. Rinne, unpublished data.

This study was fully computational and was based only on the sequence data of selected GPCRs. The data were curated in order to remove possible pseudogenes and fragmental sequences. However, it is possible to identify pseudogenes in this kind of study since biological data are mostly absent. The data and the clustering of the receptor groups have been very conservatively maintained between different releases, the original data were retrieved in 2012, and template trees were updated in 2015, 2017 and 2018 based on the different releases of the Ensembl database. If the branching was ambiguous and unstable during updates (e.g., because of bad quality sequences), receptors were not annotated, e.g., in case of gonadotropin-releasing hormone receptors.

The gene names suggested in publication **I** were considered together with IUPHAR/HUGO (Human Genome Nomenclature Committee) and will be submitted to other nomenclature committees (MGNC, Mouse Genomic Nomenclature

Committee; RGNC, Rat Genome and Nomenclature Committee, ZNC, Zebrafish Nomenclature Committee; XGC, Xenopus Gene Nomenclature Committee; CGNC, Chicken Gene Nomenclature Consortium). For the final consideration and re-evaluation of the gene names, Ensembl releases 94 and 95 were used.

5.2 Characterisation of small-molecule orexin receptor agonists Nag26 and Yan7874 (publications II and III)

The most potent compounds Nag26 and Yan7874 from the Nagahara (Nagahara et al., 2015) and Yanagisawa (Yanagisawa, 2010) series, respectively, were selected for further evaluation and characterisation. In the case of the first published compound Yan7874, the pharmacological data were defective and available only through a patent. Thus, the assessment of the Yan7874 activity on OX_1 and OX_2 was set up. Additionally, a more detailed characterisation of Nag26 was assessed in a nearly similar manner.

5.2.1 Nag26 (publication II)

In publication **II**, OX_2 selective agonist Nag26 was characterised in recombinant CHO cells. Nag26 induced strong and robust calcium elevation in both OX_1- and OX_2-expressing cells. Nag26 reached the maximum response in both cell types but showed lower potency than in orexin-A. The selectivity of Nag26 was remarkably lower than described before (Nagahara et al., 2015), and we observed a 17.6 ± 5.4-fold selectivity on OX_2 than OX_1 in calcium elevation assay. The response of Nag26 was dose response–dependently inhibited by DORA TCS-1102, as well as almost fully inhibited by Gq inhibitor UBO-QIC (1 μM).

The effect of Nag26 on PLC and Elk-1 activation was also determined in order to confirm the apparent G_q coupling. Nag26 strongly activated PLC but did not reach the maximum response of orexin-A, especially in OX_1-expressing cells. Nag26 also showed higher selectivity on OX_2, 28.8 ± 2.7-fold over OX_1 in PLC assay. PLC activity was inhibited by TCS-1102 (10 μM) and UBO-QIC (1 μM). Elk-1 was also activated by Nag26 in both cell types, but the maximum response was reached only in OX_2-expressing cells. The involvement of upstream activators ERK1/2, G_q and orexin receptors on Elk-1 activation were verified by the specific inhibitors U0126

(3 μM), UBO-QIC (1 μM) and TCS-1102 (10 μM), respectively. Nag26 induced the activation and inhibition of AC by G_S and G_i coupling, respectively. Nag26 activated AC but, surprisingly, showed no subtype selectivity among OX_1 and OX_2. Nag26 also induced AC inhibition, also indicating G_i-coupling. Both Nag26-induced activation and inhibition of AC were inhibited by TCS-1102.

Cell viability assay was set up to study possible Nag26-induced orexin receptor–dependent cell death. At low concentrations, Nag26 increased the total cell number, but it had a negative effect on cell viability, while at high concentrations, Nag26 reduced the total cell number and increased cell death. Similar effects were also partly seen in CHO-ctrl cells, indicating a not fully orexin receptor–dependent mechanism.

The results of this study and those previously reported (Nagahara et al., 2015) differ most clearly when concerning the selectivity. Previously reported selectivity of Nag26 for OX_2 over OX_1 is 70-fold, while this study suggests that the selectivity is much lower, only 20-fold. However, when normalising the previously reported Nag26 responses to the orexin-A response, the selectivity drops to 47-fold. Additionally, Nag26 was previously described as only 23-fold less potent than orexin-A, but this study found Nag26 to be much less potent, approximately100-fold. For Nag26, the Elk-1 response curve appears to be bell-shaped, as the response decreases at higher concentrations. This might be due to cell toxicity but also possible biased agonism, which, in high concentrations, Nag26 activates an alternative pathway that reduces the Elk-1 response. Nag26 also activates (G_S) and weakly inhibits (G_i) AC, and a similar blunting effect as in the Elk-1 response might be seen also in AC activity/inhibition responses. Thus, these results conclude that Nag26 is the first full small-molecule orexin receptor agonist, with possible biased agonism activity.

5.2.2 Yan7874 (publications III)

In publication **III**, the effect of Yan7874 on G_q-, G_S- and G_i-mediated activation of OX_1 and OX_2 was studied in recombinant CHO cells. For G_q-mediated activation, the effect on calcium elevation and PLC activation was determined. Yan7874 showed a strong, concentration-dependent Ca^{2+} elevation, but saturation was not reached due to low solubility. The presence of DORA TCS-1102 (10 μM)

resulted in only a slightly right-shifted dose-response curve. Additionally, Yan7874 showed only modest but concentration-dependent activation of PLC. A longer incubation time with Yan7874 resulted in a stronger response but with a bell-shaped concentration curve, indicating off-target activity in high concentrations. The G_q inhibitor UBO-QIC (1 μM) and TCS-1102 (10 μM) only partly inhibited Yan7874-induced PLC activity. The effect of Yan7874 on CHO-ctrl cells was determined in Ca^{2+} and PLC assays. While orexin-A gave no response in these cells, Yan7874 showed highly similar concentration curves as in OX_1- and OX_2-expressing cells.

The Yan7874 effect on G_S-mediated AC activity was also measured. Yan7874 showed activation at 10 μM and higher concentrations in both cell types. However, without the correlation of cAMP counts to ATP+ADP (correlation may result in a false increase of cAMP counts if compounds lower ATP+ADP counts), no activity was seen but a strong reduction of ATP+ADP counts at concentration 10 μM and above. Similarly, no significant Yan7874-induced G_i-mediated AC inhibition was seen in either of the cell types, but again, there was a strong reduction in ATP+ADP counts at the same concentrations as in the AC activity assay. Thus, coupling with G_S or G_i was not observed, and the effect of Yan7874 on ATP+ADP counts (and false AC activity) likely resulted from mitochondrial uncoupling.

These results suggested the cytotoxicity of Yan7874. Thus, the effect of Yan7874 on cell viability was determined qualitatively in stained (Hoechst 33342, PI and FDA) and non-stained cells, and quantitatively in stained cells. Apparent and rapid morphological changes in CHO-OX_1 and CHO-OX_2 cells (e.g., round shape, detachment) were observed. A similar effect on CHO-ctrl cells (and neuro-2a mouse neuroblastoma cells) was observed, and no inhibiting effect on the process with TCS-1102 in OX_1- and OX_2-expressing cells was seen, indicating that the cytotoxic mechanism of Yan7874 is not orexin receptor–dependent (or CHO cell–dependent). Additionally, data suggested that Yan7874 induces necrotic cell death since no apoptotic nuclei (condensed and fragmental) were observed.

Yan7874 showed weak and partial agonism on orexin receptors. Additionally, Yan7874 showed a rather high off-target activity and orexin receptor–independent cytotoxicity, yet is not a promising agonist for further studies. The relation between cytotoxicity and strong Ca^{2+} response was left undiscovered but might be due to rapid Ca^{2+} elevation because of the increased membrane permeability or possible mitochondrial toxicity. Additional studies of Yan7874 derivatives might be useful in

the assessment of structure-based cytotoxicity, and further in the development of non-toxic orexin receptor agonists.

5.3 Investigation of the *C. intestinalis* orexin system in vitro (publication IV)

5.3.1 Structure of the *C. intestinalis* putative orexin receptor and cognate ligand

A putative *C. intestinalis* orexin receptor, CiOX, has been previously suggested by the analysis of GPCRs in its genome (Kamesh et al., 2008), and we corroborate this finding in publication **I** of this thesis. Before molecular biology studies, we constructed a homology model of CiOX in order to identify the essential differences in the binding site and human receptor. As a result, only 7 of 24 amino acid residues, which are in close proximity of orexin-A in suggested binding mode (Karhu et al., 2015), differ between CiOX and human OX_2 binding sites: OX_2/CiOX V/T$^{3.36}$, V/S$^{3.40}$, V/S^{ECL2}, D/G ^{ECL2}, F/Y$^{5.42}$, I/A$^{6.51}$ and N/D$^{6.55}$. The differences are clustered only on one side of the pocket, relatively deep in the binding cavity, in TM3, TM5 and TM6, and one in ECL2. With the exceptions of D/G ^{ECL2} and I/A$^{6.51}$, a large steric change, other modifications are quite conservative but tend to place more polar amino acids in the ascidian receptor. Accurate modelling to explain the coevolution of receptor and ligand is made very speculative because the actual binding mode of orexin-A is not known and was thus not attempted.

Discovering a hit for the cognate peptide has proven to be very challenging. For publication **IV**, we aimed to find an open reading frame (ORF, contains a valid start and stop codon and sufficient length – more than 100 amino acids in our case) that contained fingerprints characteristic of the orexin-A peptide: the cysteine cap (pattern *CCxxxxCxC)* and the cleavage site *(ILTL/GKR)*. Furthermore, the ORF should not be part of a predicted gene, which happened a lot in the searches, and the ORF should allow itself to be retrieved in a cross-BLAST experiment (i.e., controlling that the sequence is able to retrieve the query). Diverse sequence queries (human, zebrafish, suggested invertebrate *Chordata* sequences (Jekely, 2013; Mirabeau and Joly, 2013)) were tried, as well as full-sequence or processed genes, multiple types of BLASTs (predicted *Ciona* proteins, cDNA, translated DNA), and several databases, as well as many other parameters (number of gaps in BLAST, etc.).

Yet the process was unsuccessful until late 2018. This is probably related to the presence of a relevant ORF in databases at the time, since neither the studies of Mirabeau and Joly (2013) or Jékely (2013), which used specific tailored methods for finding short peptides, do not suggest any peptide for *Tunicata*.

We finally identified a hit of putative *C. intestinalis* prepro-orexin (CiPPO), an orphan gene with XM_026835150 (NCBI; reading frame 2). The hit shared a 19 %ID and 26.6% sequence similarity with human prepro-orexin (similarity measured using EMBOSS Needle). When only the segments corresponding to orexin-A were compared, %ID raised to 22.2% and similarity to 31.1%.

To control the sequence structure relationships in the hit, a 3D structure of *C. intestinalis* putative orexin-A (Ci-orexin-A) was modelled using the solution structure of orexin-A (PDB: 1R02) (Kim et al., 2004) as a template (Figure 14). The conserved structure (among many vertebrates) of cysteine bridges in the N-terminus of orexin-A is partly different in Ci-orexin-A, resembling more orexin-A-peptide in ray-finned fishes (Xu and Volkoff, 2007). If two cysteine bridges are formed, as suggested by the model, the loops between the bridges are larger in Ci-orexin-A than in human orexin-A. Two factors support the putative peptide to be genuine. Firstly, the structural similarities of orexin-A and Ci-orexin-A lie on the same side of the peptide and might indicate higher evolutionary pressure on the conserved side of the peptide. This is a very good indicator that this part of the sequence fold is an α-helix. Similarly, the N-terminus of orexin-A is more variable between species in general, and, e.g., the primary structure of orexin-A in *D. rerio* possesses a longer N-terminus and a similar cysteine placement than Ci-orexin-A. Secondly, the model of the N-cap (containing two disulphide bridges) shows that there is space to accommodate the long inserts without straining the structure.

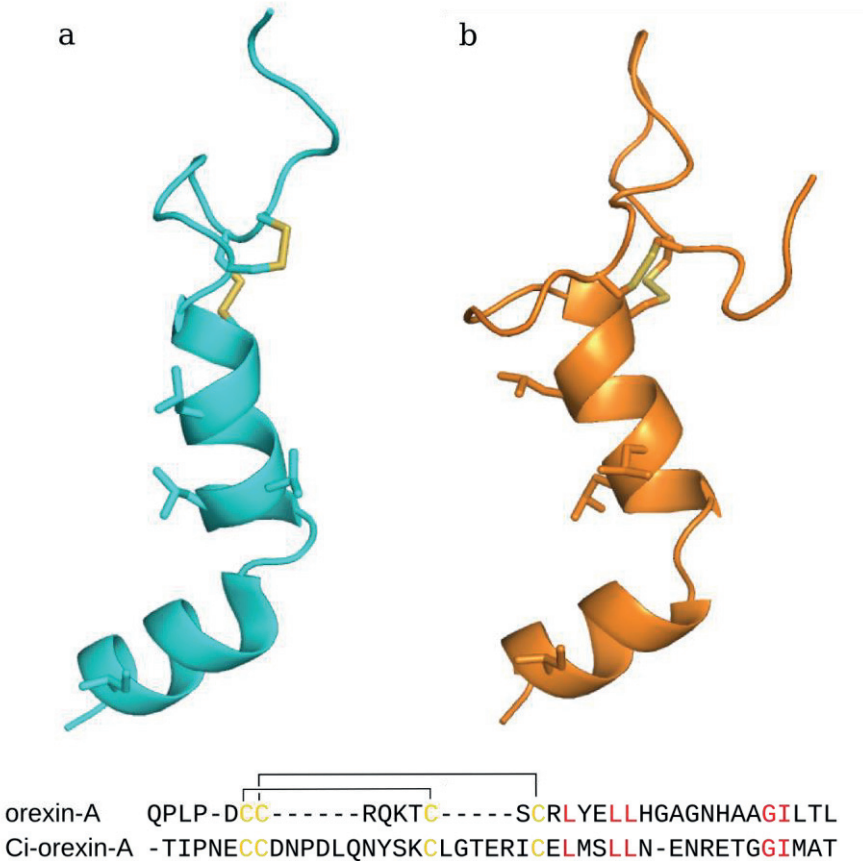

Figure 14. Above, the three-dimensional structure of human orexin-A (a; cyan; PDB: 1R02) (Kim et al., 2004) *used as a template for the homology model of suggested C. intestinalis orexin peptide (b; orange). Cysteine bridges presented in yellow, conservative amino acid residues in sticks. Below, the alignment of human orexin-A and Ci-orexin-A. Cysteines in yellow, conserved amino acids in red. Cysteine bridges represented by brackets.*

5.3.2 Verifying the activity of *C. intestinalis* putative orexin receptor and cognate ligand

To investigate the possible functionality of CiOX, the stable cell line expressing CiOX-GFP in Flp-In™ T-REx™ 293 system was generated. Surprisingly, in this cell line, Ca^{2+} elevation was observed when cells were stimulated with orexin-A, orexin-B, [Ala11,D-Leu15]-orexin-B and orexin-A_{15-33} (Table 6). Orexin-A was

slightly more potent than others. However, the maximum response with each ligand remained only at approximately 30% of the full response induced by ATP (100 μM). Small-molecule agonists Yan7874 and Nag26 did not induce Ca^{2+} elevation, but Nag26 hindered the orexin-A-induced response, similarly to DORAs TCS-1102, almorexant and the G_q inhibitor UBO-QIC, which blocked the response almost fully.

Table 6. Activation of CiOX and OX_1 in the calcium elevation (and gene reporter NFAT, CRE and serum response element (SRE)) assay (in scale +++, - presents no activation, quantitatively assessed), and the effect of inhibitors in the calcium elevation assay (in scale ///, - presents no effect, quantitatively assessed). Results from transiently transfected constructs (or chimeric CiOX-OX_1-GFP) are not discussed in this thesis.

Trans-fection*	Constructs	Expression	Ca^{2+}	Blockade of Ca^{2+}	NFAT	CRE	SRE
transient	OX_1-GFP	+ (GFP)	+++[a]				
transient	CiOX-GFP	+ (GFP)	+[d] -[a,b,c,e,f,g]				
transient	flag-OX_1	+ (flag ab)			++[d]	++[d]	-
transient	flag-CiOX	+ (flag ab)			-	-	-
stable, inducible	OX_1-GFP	+ (GFP)	+++[a,b,c,d]	///[1,2,3]			
stable, inducible	CiOX-GFP	+ (GFP)	++[a,b,c,d] -[e,f,g]	///[1,2,3]			
stable, inducible	CiOX-OX_1-GFP	+ (GFP)	-				

Agonists are as follows: orexin-A[a], orexin-B[b], [Ala[11],D-Leu[15]]-orexin-B[c], orexin-A_{15-33}[d], Nag26[e], Yan7874[f], Ci-orexin-A[g]. Inhibitors are as follows: TCS-1102[1], almorexant[2], UBO-QIC[3], Ci-orexin-A[4]. *Vectors depend on the purpose of transfection (transient/stable/inducible). Grey filling indicates data not available.

Table 7. The localisation of TAMRA-orexin-A on the plasma membrane (in scale +++, qualitatively assessed), and prevention or displacement effect of small-molecule orexin antagonists and Ci-orexin-A on the binding of TAMRA-orexin-A (in scale ///, qualitatively assessed). Cells were pre-incubated with Ci-orexin-A (prevention) before TAMRA-orexin-A was added, or Ci-orexin-A was added after TAMRA-orexin-A incubation (displacement).

Transfection	Construct	TAMRA-orexin-A	Effect of inhibitors	Effect of Ci-orexin-A (prevention)	Effect of Ci-orexin-A (displacement)
stable, inducible	OX_1-GFP	+++	///[1]	///	-
stable, inducible	CiOX-GFP	+++	-[1,2]	///	//

Inhibitors are as follow: TCS-1102[1], almorexant[2]

The activity of the identified peptide Ci-orexin-A in the calcium elevation assay on CiOX cells was determined (Table 6). Ci-orexin-A did not induce Ca^{2+} elevation in CiOX-expressing cells (or in OX_1- or OX_2-expressing cells) but potentiated the Ca^{2+} elevation induced by orexin-A (100 nM) by 2-fold at 1 µM concertation in CiOX cells. Since no apparent activation was seen, we further studied the effect of Ci-orexin-A on TAMRA-orexin-A localisation on membrane-expressing CiOX (Table 7). The co-localisation of TAMRA-orexin-A to the plasma membrane was observed in CiOX-expressing cells, as well as the internalisation with longer incubation. The pre-incubation with Ci-orexin-A blocked the co-localisation of TAMRA-orexin-A, indicating the binding of the Ci-orexin-A to CiOX. Surprisingly, small-molecule DORAs TCS-1102 (10 µM) and Almorexant (10 µM) did not affect TAMRA-orexin-A membrane co-localisation.

These results indicate that the suggested CiOX is an orexin receptor that at least partly couples to G_q when activated by the human peptides orexin-A, orexin-B, [Ala11,D-Leu15]-orexin-B and orexin-A_{15-33}. Additionally, the endogenous ligand of CiOX was suggested and validated by structural modelling and putative binding. However, the biological function of the orexin system in *C. intestinalis* remains unsolved since no *in vivo* data are available.

6 Conclusions and Perspectives

The data comprised in this thesis provide the first large-scale classification of rhodopsin-like GPCRs without human orthologues *at the level of the subtypes*. This study revealed a pool of yet-to-be-studied GPCRs, which represents work for many studies for years to come. However, understanding the function of these receptors would possibly improve the overall understanding of the function of many systems and can be exploited in many fields of science. Additionally, mappings based on sequence identity can be beneficial in identifying the evolutionary origin of many GPCRs as well as finding the best template in homology modelling studies. The data also reflect the conservative primary structure of rhodopsin family GPCRs through species. Sequence-based identification of the orphan *C. intestinalis* orexin receptor, which was further paired with a ligand in this thesis, gives additional support for the results and analysis of publication **I**. Furthermore, this thesis advocates for the quality of the Ensembl database, which has been surprisingly stable between releases during 2012–2018, even if the available genome number has more than doubled. This study also paved the way for many experiments, particularly gaining the pharmacological function of these receptors, not to mention the cartography of the B-F families in non-human vertebrates, the study of the olfactory receptors, and the mapping of invertebrate or early vertebrate receptors to human.

The near absence of available potent small-molecule orexin receptor agonists at the start of the thesis demonstrates the difficulties in discovering small-molecule ligands binding to peptide receptors. The characterised small-molecule agonists Nag26 and Yan7874 might give essential information about the binding modes of agonists in orexin receptors. The study of Yan7874, in particular, was essential to the pharmacophore modelling conducted in the laboratory (Turku et al., 2016). However, much effortful development must be done before obtaining potent orexin receptor agonists with desirable drug-like properties. Despite the relatively high affinity of Nag26, its poor water solubility dampens its potential as a drug, while the derivative YNT-185 is not lipophilic enough to cross the blood–brain barrier (Irukayama-Tomobe et al., 2017). Yan7874 shows even more poor drug-like properties; it is only a weak and partial orexin receptor agonist with orexin-independent cytotoxic effects.

The pairing of the *C. intestinalis* orexin receptor and identifying the putative prepro-orexin and orexin peptide gives a glimpse into the evolution of the orexin system. The results presented are the most successful so far and will compose a full study that includes other results: we have first identified an open reading frame for

an orphan peptide that shows characteristics of the orexin peptide, and using homology modelling, we show compatibility with the 3D structure. Following early experiments with transient expression that did not elicit signalling, we decided to use homology modelling and sequence analysis to create a chimeric humanised receptor with swapped intracellular loops, but this also did not signal through calcium (not shown in the thesis). Thus, at the moment, we have shown that the *C. intestinalis* receptor can be activated by human orexin peptides (but with low efficacy), which show that it can rely on human G proteins for signalling. Furthermore, we demonstrate binding an orphan peptide to the receptor, i.e., the ability of the peptide to block fluorescent TAMRA-orexin-A. Existing small-molecule agonists did not show activity, but tested antagonists (G_q or orexin receptor inhibitors) blocked the human peptide–induced activation in the calcium assay. However, the orphan peptide does not induce calcium elevation, which can have multiple causes, for example, poor expression and folding, signalling by pathways other than calcium, or the need of dimerisation and receptor cross-talk for signalling. Preliminary experiments carried out in Chris Langmead's group (Monash University) indeed also failed to identify a pathway using a gene reporter assay. Investigations of these issues, replicating successful experiments and designing new ones have been continued, and preliminary results have been obtained. Additionally, we have also identified putative *C. intestinalis* orexin-B peptide (not shown in the thesis), and we will rigorously test this predicted peptide.

Overall, the original function of the orexin system in tunicates remains unclear. However, the regulator of arousal is suggested, particularly in the larval stage. Despite the relatively high sequence divergence between human OX_2 and *C. intestinalis* CiOX, the binding cavities are highly similar. It would thus be extremely interesting to study coevolution at the 3D level, particularly across a range of vertebrate taxa. This should allow the identification of the binding location of the variable region of the peptide (see Figure 14) and thus provide valuable insight into the peptide binding mode, which has been elusive thus far (Karhu et al., 2015, 2019). Solving the function of the orexin system in tunicates would require additional *in vivo* experiments and could clarify the original function of the orexin system and provide valuable insights into the human orexin system.

7 References

Abascal, F., Juan, D., Jungreis, I., Martinez, L., Rigau, M., Rodriguez, J.M., et al. (2018). Loose ends: almost one in five human genes still have unresolved coding status. Nucleic Acids Res. *46*: 7070–7084.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., et al. (2000). The Genome Sequence of Drosophila melanogaster. Science (80-.). *287*: 2185–2195.

Aittaleb, M., Boguth, C.A., and Tesmer, J.J.G. (2010). Structure and Function of Heterotrimeric G Protein-Regulated Rho Guanine Nucleotide Exchange Factors. Mol. Pharmacol. *77*: 111–125.

Alexander, S.P.H., Benson, H.E., Faccenda, E., Pawson, A.J., Sharman, J.L., Spedding, M., et al. (2013). The Concise Guide to PHARMACOLOGY 2013/14: G protein-coupled receptors. Br. J. Pharmacol. *170*: 1459–1581.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*: 403–410.

Alzugaray, M.E., Adami, M.L., Diambra, L.A., Hernandez-Martinez, S., Damborenea, C., Noriega, F.G., et al. (2013). Allatotropin: An Ancestral Myotropic Neuropeptide Involved in Feeding. PLoS One *8*:.

Ammoun, S., Johansson, L., Ekholm, M.E., Holmqvist, T., Danis, A.S., Korhonen, L., et al. (2006). OX1 Orexin Receptors Activate Extracellular Signal-Regulated Kinase in Chinese Hamster Ovary Cells via Multiple Mechanisms: The Role of Ca2+ Influx in OX1 Receptor Signaling. Mol. Endocrinol. *20*: 80–99.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-M., Dehal, P., et al. (2002). Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes. Science (80-.). *297*: 1301–1310.

Asahi, S., Egashira, S.-I., Matsuda, M., Iwaasa, H., Kanatani, A., Ohkubo, M., et al. (2003). Development of an orexin-2 receptor selective agonist,[Ala11, d-Leu15] orexin-B. Bioorg. Med. Chem. Lett. *13*: 111–113.

Attwood, T.K., and Findlay, J.B. (1994). Fingerprinting G-protein-coupled receptors. Protein Eng. *7*: 195–203.

Baimel, C., and Borgland, S.L. (2012). Hypocretin modulation of drug-induced synaptic plasticity. Prog. Brain Res. *198*: 123–131.

Ballesteros, J.A., and Weinstein, H. (1995). [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods Neurosci. *25*: 366–428.

Bargmann, C.I. (1998). Neurobiology of the Caenorhabditis elegans genome.

Science *282*: 2028–33.

Bauknecht, P., and Jékely, G. (2017). Ancient coexistence of norepinephrine, tyramine, and octopamine signaling in bilaterians. BMC Biol. *15*: 6.

Beuckmann, C.T., Sinton, C.M., Williams, S.C., Richardson, J.A., Hammer, R.E., Sakurai, T., et al. (2004). Expression of a Poly-Glutamine-Ataxin-3 Transgene in Orexin Neurons Induces Narcolepsy–Cataplexy in the Rat. J. Neurosci. *24*: 4469 LP – 4477.

Beuckmann, C.T., Suzuki, M., Ueno, T., Nagaoka, K., Arai, T., and Higashiyama, H. (2017). In Vitro and In Silico Characterization of Lemborexant (E2006), a Novel Dual Orexin Receptor Antagonist. J. Pharmacol. Exp. Ther. *362*: 287 LP – 295.

Bjarnadóttir, T.K., Fredriksson, R., Höglund, P.J., Gloriam, D.E., Lagerström, M.C., and Schiöth, H.B. (2004). The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. Genomics *84*: 23–33.

Bockaert, J., and Pin, J.P. (1999). Molecular tinkering of G protein-coupled receptors: an evolutionary success. EMBO J. *18*: 1723–1729.

Bonaventure, P., Shelton, J., Yun, S., Nepomuceno, D., Sutton, S., Aluisio, L., et al. (2015). Characterization of JNJ-42847922, a Selective Orexin-2 Receptor Antagonist, as a Clinical Candidate for the Treatment of Insomnia. J. Pharmacol. Exp. Ther. *354*: 471 LP – 482.

Bosch, T.C.G., Klimovich, A., Domazet-Lošo, T., Gründer, S., Holstein, T.W., Jékely, G., et al. (2017). Back to the Basics: Cnidarians Start to Fire. Trends Neurosci. *40*: 92–105.

Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., et al. (2016). The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. *48*: 427–437.

Brisbare-Roch, C., Dingemanse, J., Koberstein, R., Hoever, P., Aissaoui, H., Flores, S., et al. (2007). Promotion of sleep by targeting the orexin system in rats, dogs and humans. Nat. Med. *13*: 150–155.

Brody, T., and Cravchik, A. (2000). Drosophila melanogaster G protein-coupled receptors. J. Cell Biol. *150*: F83-8.

Bulenger, S., Marullo, S., and Bouvier, M. (2005). Emerging role of homo- and heterodimerization in G-protein-coupled receptor biosynthesis and maturation. Trends Pharmacol. Sci. *26*: 131–137.

Burbach, J.P., and Meijer, O.C. (1992). The structure of neuropeptide receptors. Eur. J. Pharmacol. *227*: 1–18.

Burbach, J.P.H. (2004). [The 2004 Nobel Prize for Physiology or Medicine for research into smell receptors and the organization of the olfactory system]. Ned. Tijdschr. Geneeskd. *148*: 2576–2579.

Cano, M., Grima, P.M., and Benet, A.P. (2014). 2-(2-aminophenoxy)-3-chloronaphthalene-1,4-dione compounds having orexin 2 receptor agonist activity. Patent WO2014198880.

Caputo Barucchi, V., Giovannotti, M., Nisi Cerioni, P., and Splendiani, A. (2013). Genome Duplication in Early Vertebrates: Insights from Agnathan Cytogenetics. Cytogenet. Genome Res. *141*: 80–89.

Carpentier, M., and Chomilier, J. (2019). Protein multiple alignments: sequence-based versus structure-based programs. Bioinformatics.

Chandrashekar, I.R., Dike, A., and Cowsik, S.M. (2004). Membrane-induced structure of the mammalian tachykinin neuropeptide gamma. J.Struct.Biol. *148*: 315–325.

Chemelli, R.M., Willie, J.T., Sinton, C.M., Elmquist, J.K., Scammell, T., Lee, C., et al. (1999). Narcolepsy in orexin Knockout Mice: Molecular Genetics of Sleep Regulation. Cell *98*: 437–451.

Clark, R.B. (2013). Profile of Brian K. Kobilka and Robert J. Lefkowitz, 2012 Nobel Laureates in Chemistry. Proc. Natl. Acad. Sci. *110*: 5274 LP – 5275.

Coleman, B.D., Marivin, A., Parag-Sharma, K., DiGiacomo, V., Kim, S., Pepper, J.S., et al. (2015). Evolutionary Conservation of a GPCR-Independent Mechanism of Trimeric G Protein Activation. Mol. Biol. Evol. *33*: 820–837.

Consortium*, T.C. elegans S. (1998). Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. Science (80-.). *282*: 2012–2018.

Coutant, J., Curmi, P.A., Toma, F., and Monti, J.P. (2007). NMR Solution Structure of Neurotensin in Membrane-Mimetic Environments: Molecular Basis for Neurotensin Receptor Recognition. Biochemistry *46*: 5656–5663.

Cox, C.D., Breslin, M.J., Whitman, D.B., Schreier, J.D., McGaughey, G.B., Bogusky, M.J., et al. (2010). Discovery of the Dual Orexin Receptor Antagonist [(7R)-4-(5-Chloro-1,3-benzoxazol-2-yl)-7-methyl-1,4-diazepan-1-yl][5-methyl-2-(2H-1,2,3-triazol-2-yl)phenyl]methanone (MK-4305) for the Treatment of Insomnia. J. Med. Chem. *53*: 5320–5332.

Crow, K.D., and Wagner, G.P. (2006). What Is the Role of Genome Duplication in the Evolution of Complexity and Diversity? Mol. Biol. Evol. *23*: 887–892.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2014). Ensembl 2015. Nucleic Acids Res. *43*: D662–D669.

Daaka, Y., Luttrell, L.M., and Lefkowitz, R.J. (1997). Switching of the coupling of the β2-adrenergic receptor to different G proteins by protein kinase A. Nature *390*: 88–91.

Dahlberg, C., Auger, H., Dupont, S., Sasakura, Y., Thorndyke, M., and Joly, J.-S. (2009). Refining the Ciona intestinalis model of central nervous system

regeneration. PLoS One *4*: e4458–e4458.

Darker, J.G., Porter, R.A., Eggleston, D.S., Smart, D., Brough, S.J., Sabido-David, C., et al. (2001). Structure–activity analysis of truncated orexin-A analogues at the orexin-1 receptor. Bioorg. Med. Chem. Lett. *11*: 737–740.

Daulat, A.M., Maurice, P., and Jockers, R. (2009). Recent methodological advances in the discovery of GPCR-associated protein complexes. Trends Pharmacol. Sci. *30*: 72–78.

Davidson, F.F., Loewen, P.C., and Khorana, H.G. (1994). Structure and function in rhodopsin: replacement by alanine of cysteine residues 110 and 187, components of a conserved disulfide bond in rhodopsin, affects the light-activated metarhodopsin II state. Proc. Natl. Acad. Sci. U. S. A. *91*: 4029–33.

Dehal, P., and Boore, J.L. (2005). Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. PLoS Biol. *3*: e314.

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., Tomaso, A. De, et al. (2002). The Draft Genome of Ciona intestinalis: Insights into Chordate and Vertebrate Origins. Science (80-.). *298*: 2157–2167.

Deupi, X. (2014). Relevance of rhodopsin studies for GPCR activation. Biochim. Biophys. Acta - Bioenerg. *1837*: 674–682.

Devreotes, P.N. (1994). G protein-linked signaling pathways control the developmental program of Dictyostelium. Neuron *12*: 235–41.

DeVries, M.E., Kelvin, A.A., Xu, L., Ran, L., Robinson, J., and Kelvin, D.J. (2006). Defining the origins and evolution of the chemokine/chemokine receptor system. J. Immunol. *176*: 401–15.

Dike, A., and Cowsik, S.M. (2006). Three-dimensional structure of neuropeptide k bound to dodecylphosphocholine micelles. Biochemistry *45*: 2994–3004.

Dixon, R.A.F., Kobilka, B.K., Strader, D.J., Benovic, J.L., Dohlman, H.G., Frielle, T., et al. (1986). Cloning of the gene and cDNA for mammalian β-adrenergic receptor and homology with rhodopsin. Nature *321*: 75–79.

Dohlman, H.G., Thorner, J., Caron, M.G., and Lefkowitz, R.J. (1991). Model Systems for the Study of Seven-Transmembrane-Segment Receptors. Annu. Rev. Biochem. *60*: 653–688.

Dong, S.S., Goddard 3rd, W.A., and Abrol, R. (2016). Conformational and Thermodynamic Landscape of GPCR Activation from Theory and Computation. Biophys. J. *110*: 2618–2629.

Donoghue, P.C.J., and Purnell, M.A. (2005). Genome duplication, extinction and vertebrate evolution. Trends Ecol. Evol. *20*: 312–319.

Dores, R.M. (2011). Hagfish, Genome Duplications, and RFamide Neuropeptide Evolution. Endocrinology *152*: 4010–3.

Dugovic, C., Shelton, J.E., Aluisio, L.E., Fraser, I.C., Jiang, X., Sutton, S.W., et al. (2009). Blockade of orexin-1 receptors attenuates orexin-2 receptor antagonism-induced sleep promotion in the rat. J. Pharmacol. Exp. Ther. *330*: 142–151.

Edwards, C.M., Abusnana, S., Sunter, D., Murphy, K.G., Ghatei, M.A., and Bloom, S.R. (1999). The effect of the orexins on food intake: comparison with neuropeptide Y, melanin-concentrating hormone and galanin. J. Endocrinol. *160*: R7-12.

Eipper, B.A., Stoffers, D.A., and Mains, R.E. (1992). The Biosynthesis of Neuropeptides: Peptide alpha-Amidation. Annu. Rev. Neurosci. *15*: 57–85.

Ekholm, M.E., Johansson, L., and Kukkonen, J.P. (2007). IP3-independent signalling of OX1 orexin/hypocretin receptors to Ca2+ influx and ERK. Biochem. Biophys. Res. Commun. *353*: 475–480.

Elphick, M.R., Mirabeau, O., and Larhammar, D. (2018). Evolution of neuropeptide signalling systems. J. Exp. Biol. *221*: jeb151092.

Estabrooke, I. V, McCarthy, M.T., Ko, E., Chou, T.C., Chemelli, R.M., Yanagisawa, M., et al. (2001). Fos Expression in Orexin Neurons Varies with Behavioral State. J. Neurosci. *21*: 1656 LP – 1662.

Fabio, R. Di, Pellacani, A., Faedo, S., Roth, A., Piccoli, L., Gerrard, P., et al. (2011). Discovery process and pharmacological characterization of a novel dual orexin 1 and orexin 2 receptor antagonist useful for treatment of sleep disorders. Bioorg. Med. Chem. Lett. *21*: 5562–5567.

Fang, Y., Kenakin, T., and Liu, C. (2015). Editorial: Orphan GPCRs As Emerging Drug Targets. Front. Pharmacol. *6*: 295.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., et al. (2012). Ensembl 2012. Nucleic Acids Res. *40*: D84–D90.

Fredriksson, R., Lagerström, M.C., Lundin, L.-G., and Schiöth, H.B. (2003). The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. Mol. Pharmacol. *63*: 1256–1272.

Fredriksson, R., Larson, E.T., Yan, Y.-L., Postlethwait, J.-H., and Larhammar, D. (2004). Novel neuropeptide Y Y2-like receptor subtype in zebrafish and frogs supports early vertebrate chromosome duplications. J. Mol. Evol. *58*: 106–14.

Funato, H., Tsai, A.L., Willie, J.T., Kisanuki, Y., Williams, S.C., Sakurai, T., et al. (2009). Enhanced Orexin Receptor-2 Signaling Prevents Diet-Induced Obesity and Improves Leptin Sensitivity. Cell Metab. *9*: 64–76.

Gacasan, S.B., Baker, D.L., and Parrill, A.L. (2017). G protein-coupled receptors: the evolution of structural insight. AIMS Biophys. *4*: 491–527.

Galandrin, S., Oligny-Longpré, G., and Bouvier, M. (2007). The evasive nature of drug efficacy: implications for drug discovery. Trends Pharmacol. Sci. *28*: 423–

430.

Garcia-Fernàndez, J. (2005). The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. *6*: 881–892.

German, N.A., Decker, A.M., Gilmour, B.P., Thomas, B.F., and Zhang, Y. (2013). Truncated Orexin Peptides: Structure-Activity Relationship Studies. ACS Med. Chem. Lett. *4*: 1224–1227.

Gilman, A.G. (1987). G proteins: transducers of receptor-generated signals. Annu. Rev. Biochem. *56*: 615–649.

Girault, E.M., Yi, C.-X., Fliers, E., and Kalsbeek, A. (2012). Orexins, feeding, and energy balance. Prog. Brain Res. *198*: 47–64.

Gloriam, D.E., Fredriksson, R., and Schiöth, H.B. (2007). The G protein-coupled receptor subset of the rat genome. BMC Genomics *8*: 338.

Gloriam, D.E.I., Bjarnadóttir, T.K., Yan, Y.-L., Postlethwait, J.H., Schiöth, H.B., and Fredriksson, R. (2005). The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish. Mol. Phylogenet. Evol. *35*: 470–82.

Good, S., Yegorov, S., Martijn, J., Franck, J., and Bogerd, J. (2012). New Insights into Ligand-Receptor Pairing and Coevolution of Relaxin Family Peptides and Their Receptors in Teleosts. Int. J. Evol. Biol. *2012*: 1–14.

Gotter, A.L., Forman, M.S., Harrell, C.M., Stevens, J., Svetnik, V., Yee, K.L., et al. (2016). Orexin 2 Receptor Antagonism is Sufficient to Promote NREM and REM Sleep from Mouse to Man. Sci. Rep. *6*: 27147.

Granier, S., Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Weis, W.I., et al. (2012). Structure of the delta opioid receptor bound to naltrindole. Nature *485*: 400–404.

Gregory, S.G., Barlow, K.F., McLay, K.E., Kaul, R., Swarbreck, D., Dunham, A., et al. (2006). The DNA sequence and biological annotation of human chromosome 1. Nature *441*: 315–321.

Gregory, T.R. (2001). The Bigger the C-Value, the Larger the Cell: Genome Size and Red Blood Cell Size in Vertebrates. Blood Cells, Mol. Dis. *27*: 830–843.

Grimmelikhuijzen, C.J.P., and Hauser, F. (2012). Mini-review: The evolution of neuropeptide signaling. Regul. Pept. *177*: S6–S9.

Grimmelikhuijzen, C.J.P., Leviev, I., and Carstensen, K. (1996). Peptides in the Nervous Systems of Cnidarians: Structure, Function, and Biosynthesis. Int. Rev. Cytol. *167*: 37–89.

Gurevich, V. V, and Gurevich, E. V (2019). GPCR Signaling Regulation: The Role of GRKs and Arrestins . Front. Pharmacol. *10*: 125.

Harding, S.D., Sharman, J.L., Faccenda, E., Southan, C., Pawson, A.J., Ireland, S.,

et al. (2017). The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. Nucleic Acids Res. *46*: D1091–D1106.

Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schiöth, H.B., and Gloriam, D.E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. Nat. Rev. Drug Discov. *16*: 829–842.

Heifetz, A., Morris, G.B., Biggin, P.C., Barker, O., Fryatt, T., Bentley, J., et al. (2012). Study of Human Orexin-1 and -2 G-Protein-Coupled Receptors with Novel and Published Antagonists by Modeling, Molecular Dynamics Simulations, and Site-Directed Mutagenesis. Biochemistry *51*: 3178–3197.

Hewes, R.S., and Taghert, P.H. (2001). Neuropeptides and neuropeptide receptors in the Drosophila melanogaster genome. Genome Res. *11*: 1126–42.

Hirose, M., Egashira, S., Goto, Y., Hashihayata, T., Ohtake, N., Iwaasa, H., et al. (2003). N-acyl 6, 7-dimethoxy-1, 2, 3, 4-tetrahydroisoquinoline: the first orexin-2 receptor selective non-peptidic antagonist. Bioorg. Med. Chem. Lett. *13*: 4497–4499.

Ho, Y.-C., Lee, H.-J., Tung, L.-W., Liao, Y.-Y., Fu, S.-Y., Teng, S.-F., et al. (2011). Activation of Orexin 1 Receptors in the Periaqueductal Gray of Male Rats Leads to Antinociception via Retrograde Endocannabinoid (2-Arachidonoylglycerol)-Induced Disinhibition. J. Neurosci. *31*: 14600 LP – 14610.

Holland, P., and Goadsby, P.J. (2007). The Hypothalamic Orexinergic System: Pain and Primary Headaches. Headache J. Head Face Pain *47*: 951–962.

Holland, P.W.H. (1999). Gene duplication: Past, present and future. Semin. Cell Dev. Biol. *10*: 541–547.

Hollander, J.A., Lu, Q., Cameron, M.D., Kamenecka, T.M., and Kenny, P.J. (2008). Insular hypocretin transmission regulates nicotine reward. Proc. Natl. Acad. Sci. *105*: 19480 LP – 19485.

Holmqvist, T., Johansson, L., Östman, M., Ammoun, S., Åkerman, K.E.O., and Kukkonen, J.P. (2005). OX1 orexin receptors couple to adenylyl cyclase regulation via multiple mechanisms. J. Biol. Chem. *280*: 6570–6579.

Holst, B., and Schwartz, T.W. (2003). Molecular mechanism of agonism and inverse agonism in the melanocortin receptors: Zn(2+) as a structural and functional probe. Ann. N. Y. Acad. Sci. *994*: 1–11.

Hook, V., Funkelstein, L., Lu, D., Bark, S., Wegrzyn, J., and Hwang, S.-R. (2008). Proteases for Processing Proneuropeptides into Peptide Neurotransmitters and Hormones. Annu. Rev. Pharmacol. Toxicol. *48*: 393–423.

Horodyski, F.M., Verlinden, H., Filkin, N., Vandersmissen, H.P., Fleury, C., Reynolds, S.E., et al. (2011). Isolation and functional characterization of an allatotropin receptor from Manduca sexta. Insect Biochem. Mol. Biol. *41*: 804–814.

Hoyer, D. (2017). 5-HT Receptor Nomenclature: Naming Names, Does It Matter? A Tribute to Maurice Rapport. ACS Chem. Neurosci. *8*: 908–919.

Hoyer, D., and Bartfai, T. (2012). Neuropeptides and Neuropeptide Receptors: Drug Targets, and Peptide and Non-Peptide Ligands: a Tribute to Prof. *Dieter Seebach*. Chem. Biodivers. *9*: 2367–2387.

Huang, W., Manglik, A., Venkatakrishnan, A.J., Laeremans, T., Feinberg, E.N., Sanborn, A.L., et al. (2015). Structural insights into mu-opioid receptor activation. Nature *524*: 315–321.

Hughes, J.P., Rees, S., Kalindjian, S.B., and Philpott, K.L. (2011). Principles of early drug discovery. Br. J. Pharmacol. *162*: 1239–1249.

Hyndman, K.A., Miyamoto, M.M., and Evans, D.H. (2009). Phylogeny, taxonomy, and evolution of the endothelin receptor gene family. Mol. Phylogenet. Evol. *52*: 677–687.

Irannejad, R. (2014). GPCR signaling along the endocytic pathway. Curr. Opin. Cell Biol. *27*: 109–116.

Irukayama-Tomobe, Y., Ogawa, Y., Tominaga, H., Ishikawa, Y., Hosokawa, N., Ambai, S., et al. (2017). Nonpeptide orexin type-2 receptor agonist ameliorates narcolepsy-cataplexy symptoms in mouse models. Proc. Natl. Acad. Sci. *114*: 5731 LP – 5736.

Irvine, S.W. laboratory at U. (2019). Membrane Proteins of Known 3D Structures.

Isberg, V., Vroling, B., Kant, R. van der, Li, K., Vriend, G., and Gloriam, D. (2014). GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res. *42*: D422–D425.

Jacobson, L.H., Callander, G.E., and Hoyer, D. (2014). Suvorexant for the treatment of insomnia. Expert Rev. Clin. Pharmacol. *7*: 711–730.

Janes, R.W., Peapus, D.H., and Wallace, B.A. (1994). The crystal structure of human endothelin. Nat.Struct.Mol.Biol. *1*: 311–319.

Janto, K., Prichard, J.R., and Pusalavidyasagar, S. (2018). An Update on Dual Orexin Receptor Antagonists and Their Potential Role in Insomnia Therapeutics. J. Clin. Sleep Med. *14*: 1399–1408.

Jäntti, M.H., Putula, J., Somerharju, P., Frohman, M.A., and Kukkonen, J.P. (2012). OX1 orexin/hypocretin receptor activation of phospholipase D. Br. J. Pharmacol. *165*: 1109–1123.

Jäntti, M.H., Putula, J., Turunen, P.M., Näsman, J., Reijonen, S., Lindqvist, C., et al. (2013). Autocrine Endocannabinoid Signaling through CB₁ Receptors Potentiates OX₁ Orexin Receptor Signaling. Mol. Pharmacol. *83*: 621 LP – 632.

Jeffery, W.R. (2018). Regeneration and Aging in the Tunicate Ciona intestinalis.

Conn's Handb. Model. Hum. Aging 521–531.

Jekely, G. (2013). Global view of the evolution and diversity of metazoan neuropeptide signaling. Proc. Natl. Acad. Sci. *110*: 8702–8707.

Johansson, L., Ekholm, M.E., and Kukkonen, J.P. (2008). Multiple phospholipase activation by OX1 orexin/hypocretin receptors. Cell. Mol. Life Sci. *65*: 1948–1956.

Jöhren, O., Neidert, S.J., Kummer, M., Dendorfer, A., and Dominiak, P. (2001). Prepro-Orexin and Orexin Receptor mRNAs Are Differentially Expressed in Peripheral Tissues of Male and Female Rats. Endocrinology *142*: 3324–3331.

Kamano, H., Burk, B., Noben-Trauth, K., and Klempnauer, K.-H. (1995). Differential splicing of the mouse B-myb gene. Oncogene *11*: 2575–2582.

Kamesh, N., Aradhyam, G.K., and Manoj, N. (2008). The repertoire of G protein-coupled receptors in the sea squirt Ciona intestinalis. BMC Evol. Biol. *8*: 129.

Karhu, L., Magarkar, A., Bunker, A., and Xhaard, H. (2019). Determinants of Orexin Receptor Binding and Activation—A Molecular Dynamics Study. J. Phys. Chem. B *123*: 2609–2622.

Karhu, L., Turku, A., and Xhaard, H. (2015). Modeling of the OX1R–orexin-A complex suggests two alternative binding modes. BMC Struct. Biol. *15*: 9.

Karteris, E., Machado, R.J., Chen, J., Zervou, S., Hillhouse, E.W., and Randeva, H.S. (2005). Food deprivation differentially modulates orexin receptor expression and signaling in rat hypothalamus and adrenal cortex. Am. J. Physiol. Metab. *288*: E1089–E1100.

Karteris, E., Randeva, H.S., Grammatopoulos, D.K., Jaffe, R.B., and Hillhouse, E.W. (2001). Expression and Coupling Characteristics of the CRH and Orexin Type 2 Receptors in Human Fetal Adrenals. J. Clin. Endocrinol. Metab. *86*: 4512–4519.

Kaslin, J., Nystedt, J.M., Östergård, M., Peitsaro, N., and Panula, P. (2004). The Orexin/Hypocretin System in Zebrafish Is Connected to the Aminergic and Cholinergic Systems. J. Neurosci. *24*: 2678 LP – 2689.

Kataoka, H., Toschi, A., Li, J.P., Carney, R.L., Schooley, D.A., and Kramer, S.J. (1989). Identification of an Allatotropin from Adult Manduca Sexta Science (80-.). *243*: 1481 LP – 1483.

Katoh, H., Aoki, J., Yamaguchi, Y., Kitano, Y., Ichikawa, A., and Negishi, M. (1998). Constitutively active Gα12, Gα13, and Gαq induce Rho-dependent neurite retraction through different signaling pathways. J. Biol. Chem. *273*: 28700–28707.

Kawada, T., Ogasawara, M., Sekiguchi, T., Aoyama, M., Hotta, K., Oka, K., et al. (2011). Peptidomic Analysis of the Central Nervous System of the Protochordate, Ciona intestinalis: Homologs and Prototypes of Vertebrate Peptides and Novel Peptides. Endocrinology *152*: 2416–2427.

Kawada, T., Sekiguchi, T., Sakai, T., Aoyama, M., and Satake, H. (2010).

Neuropeptides, Hormone Peptides, and Their Receptors in *Ciona intestinalis:* An Update. Zoolog. Sci. *27*: 134–153.

Kenakin, T. (1996). Receptor conformational induction versus selection: all part of the same energy landscape: Agonists can differentially stabilize multiple active states of receptors. Trends Pharmacol. Sci. *17*: 190–191.

Kilduff, T.S., and Peyron, C. (2000). The hypocretin/orexin ligand–receptor system: implications for sleep and sleep disorders. Trends Neurosci. *23*: 359–365.

Kim, H.-Y., Hong, E.-M., Kim, J., and Lee, W.-T. (2004). Solution structure of human orexin-A: regulator of appetite and wakefulness. BMB Rep. *37*: 565–573.

Koehbach, J., O'Brien, M., Muttenthaler, M., Miazzo, M., Akcan, M., Elliott, A.G., et al. (2013). Oxytocic plant cyclotides as templates for peptide G protein-coupled receptor ligand design. Proc.Natl.Acad.Sci.USA *110*: 21183–21188.

Kolakowski, L.F. (1994). GCRDb: a G-protein-coupled receptor database. Receptors Channels *2*: 1–7.

Korotkova, T.M., Sergeeva, O.A., Eriksson, K.S., Haas, H.L., and Brown, R.E. (2003). Excitation of Ventral Tegmental Area Dopaminergic and Nondopaminergic Neurons by Orexins/Hypocretins. J. Neurosci. *23*: 7 LP – 11.

Kortenjann, M., Thomae, O., and Shaw, P.E. (1994). Inhibition of v-raf-dependent c-fos expression and transformation by a kinase-defective mutant of the mitogen-activated protein kinase Erk2. Mol. Cell. Biol. *14*: 4815 LP – 4824.

Krishnan, A. (2015). Evolution of the G protein-coupled receptor signaling system: Genomic and phylogenetic analyses.

Krishnan, A., Almén, M.S., Fredriksson, R., and Schiöth, H.B. (2013). Remarkable similarities between the hemichordate (Saccoglossus kowalevskii) and vertebrate GPCR repertoire. Gene *526*: 122–33.

Kukkonen, J.P. (2016). OX2 orexin/hypocretin receptor signal transduction in recombinant Chinese hamster ovary cells. Cell. Signal. *28*: 51–60.

Kukkonen, J.P., and Åkerman, K.E.O. (2001). Orexin receptors couple to Ca2+ channels different from store-operated Ca2+ channels. Neuroreport *12*: 2017–2020.

Kukkonen, J.P., and Leonard, C.S. (2014). Orexin/hypocretin receptor signalling cascades. Br. J. Pharmacol. *171*: 314–331.

Kuriyama, A., and Tabata, H. (2017). Suvorexant for the treatment of primary insomnia: A systematic review and meta-analysis. Sleep Med. Rev. *35*: 1–7.

Lagerström, M.C., Hellström, A.R., Gloriam, D.E., Larsson, T.P., Schiöth, H.B., and Fredriksson, R. (2006). The G protein-coupled receptor subset of the chicken genome. PLoS Comput. Biol. *2*: e54.

Lagerström, M.C., and Schiöth, H.B. (2008). Structural diversity of G protein-

coupled receptors and significance for drug discovery. Nat. Rev. Drug Discov. *7*: 339–57.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*: 860–921.

Lang, M., Söll, R.M., Dürrenberger, F., Dautzenberg, F.M., and Beck-Sickinger, A.G. (2004). Structure− activity studies of orexin A and orexin B at the human orexin 1 and orexin 2 receptors led to orexin 2 receptor selective and orexin 1 receptor preferring ligands. J. Med. Chem. *47*: 1153–1160.

Langmead, C.J., Jerman, J.C., Brough, S.J., Scott, C., Porter, R.A., and Herdon, H.J. (2004). Characterisation of the binding of [3H]-SB-674042, a novel nonpeptide antagonist, to the human orexin-1 receptor. Br. J. Pharmacol. *141*: 340–346.

Larhammar, D., and Salaneck, E. (2004). Molecular evolution of NPY receptor subtypes. Neuropeptides *38*: 141–151.

Larhammar, D., Xu, B., and Bergqvist, C.A. (2014). Unexpected multiplicity of QRFP receptors in early vertebrate evolution. Front. Neurosci. *8*: 337.

Lawler, O.A., Miggin, S.M., and Kinsella, B.T. (2001). Protein Kinase A-mediated Phosphorylation of Serine 357 of the Mouse Prostacyclin Receptor Regulates Its Coupling to G_s -, to G_i -, and to G_q -coupled Effector Signaling. J. Biol. Chem. *276*: 33596–33607.

Lawrence, A.J., Cowen, M.S., Yang, H.-J., Chen, F., and Oldfield, B. (2006). The orexin system regulates alcohol-seeking in rats. Br. J. Pharmacol. *148*: 752–759.

Lecea, L. de, Kilduff, T.S., Peyron, C., Gao, X.-B., Foye, P.E., Danielson, P.E., et al. (1998). The hypocretins: Hypothalamus-specific peptides with neuroexcitatory activity. Proc. Natl. Acad. Sci. *95*: 322–327.

Lee, J., Bang, E., Chae, K., Kim, J., Lee, D.W., and Lee, W. (1999). Solution structure of a new hypothalamic neuropeptide, human hypocretin-2/orexin-B. Eur. J. Biochem. *266*: 831–839.

Lefkowitz, R.J. (1994). Rodbell and Gilman win 1994 Nobel Prize for Physiology and Medicine. Trends Pharmacol. Sci. *15*: 442–444.

Leino, T.O., Turku, A., Yli-Kauhaluoma, J., Kukkonen, J.P., Xhaard, H., and Wallén, E.A.A. (2018). Azulene-based compounds for targeting orexin receptors. Eur. J. Med. Chem. *157*: 88–100.

Leonard, C.S., and Kukkonen, J.P. (2014). Orexin/hypocretin receptor signalling: a functional perspective. Br. J. Pharmacol. *171*: 294–313.

Letavic, M.A., Bonaventure, P., Carruthers, N.I., Dugovic, C., Koudriakova, T., Lord, B., et al. (2015). Novel Octahydropyrrolo[3,4-c]pyrroles Are Selective Orexin-2 Antagonists: SAR Leading to a Clinical Candidate. J. Med. Chem. *58*: 5620–5636.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*: 147–151.

Li, C., and Kim, K. (2010). Neuropeptide gene families in Caenorhabditis elegans. Adv. Exp. Med. Biol. *692*: 98–137.

Li, Q., Tachie-Baffour, Y., Liu, Z., Baldwin, M.W., Kruse, A.C., and Liberles, S.D. (2015). Non-classical amine recognition evolved in a large clade of olfactory receptors. Elife *4*: e10441.

Lin, L., Faraco, J., Li, R., Kadotani, H., Rogers, W., Lin, X., et al. (1999). The Sleep Disorder Canine Narcolepsy Is Caused by a Mutation in the Hypocretin (Orexin) Receptor 2 Gene. Cell *98*: 365–376.

Lund, P.-E., Shariatmadari, R., Uustare, A., Detheux, M., Parmentier, M., Kukkonen, J.P., et al. (2000). The orexin OX1 receptor activates a novel Ca2+ influx pathway necessary for coupling to phospholipase C. J. Biol. Chem. *275*: 30806–30812.

Malendowicz, L.K., Tortorella, C., and Nussdorfer, G.G. (1999). Orexins stimulate corticosterone secretion of rat adrenocortical cells, through the activation of the adenylate cyclase-dependent signaling cascade. J. Steroid Biochem. Mol. Biol. *70*: 185–188.

Malherbe, P., Borroni, E., Gobbi, L., Knust, H., Nettekoven, M., Pinard, E., et al. (2009). Biochemical and behavioural characterization of EMPA, a novel high-affinity, selective antagonist for the OX2 receptor. Br. J. Pharmacol. *156*: 1326–1341.

Malherbe, P., Roche, O., Marcuz, A., Kratzeisen, C., Wettstein, J.G., and Bissantz, C. (2010). Mapping the Binding Pocket of Dual Antagonist Almorexant to Human Orexin 1 and Orexin 2 Receptors: Comparison with the Selective OX₁ Antagonist SB-674042 and the Selective OX₂ Antagonist N Mol. Pharmacol. *78*: 81 LP – 93.

Malnic, B., Godfrey, P.A., and Buck, L.B. (2004). The human olfactory receptor gene family. Proc. Natl. Acad. Sci. U. S. A. *101*: 2584 LP – 2589.

Manglik, A., and Kruse, A.C. (2017). Structural Basis for G Protein-Coupled Receptor Activation. Biochemistry *56*: 5628–5634.

Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Mathiesen, J.M., Sunahara, R.K., et al. (2012). Crystal structure of the {mu}-opioid receptor bound to a morphinan antagonist. Nature *485*: 321–326.

Mantha, A.K., Chandrashekar, I.R., Baquer, N.Z., and Cowsik, S.M. (2004). Three dimensional structure of Mammalian tachykinin Peptide neurokinin B bound to lipid micelles. J.Biomol.Struct.Dyn. *22*: 137–148.

Marcus, J.N., Aschkenasi, C.J., Lee, C.E., Chemelli, R.M., Saper, C.B.,

Yanagisawa, M., et al. (2001). Differential expression of orexin receptors 1 and 2 in the rat brain. J. Comp. Neurol. *435*: 6–25.

Maurice, P., Guillaume, J.-L., Benleulmi-Chaachoua, A., Daulat, A.M., Kamal, M., and Jockers, R. (2011). GPCR-Interacting Proteins, Major Players of GPCR Function. Adv. Pharmacol. *62*: 349–380.

McAtee, L.C., Sutton, S.W., Rudolph, D.A., Li, X., Aluisio, L.E., Phuong, V.K., et al. (2004). Novel substituted 4-phenyl-[1,3]dioxanes: potent and selective orexin receptor 2 (OX2R) antagonists. Bioorg. Med. Chem. Lett. *14*: 4225–4229.

McElhinny, C.J., Lewin, A.H., Mascarella, S.W., Runyon, S., Brieaddy, L., and Carroll, F.I. (2012). Hydrolytic instability of the important orexin 1 receptor antagonist SB-334867: Possible confounding effects on in vivo and in vitro studies. Bioorg. Med. Chem. Lett. *22*: 6661–6664.

Mcknight, A.J., and Gordon, S. (1996). EGF-TM7: a novel subfamily of seven-transmembrane-region leukocyte cell-surface molecules. Immunol. Today *17*: 283–287.

Mieda, M., Hasegawa, E., Kisanuki, Y.Y., Sinton, C.M., Yanagisawa, M., and Sakurai, T. (2011). Differential Roles of Orexin Receptor-1 and -2 in the Regulation of Non-REM and REM Sleep. J. Neurosci. *31*: 6518 LP – 6526.

Milasta, S., Evans, N.A., Ormiston, L., Wilson, S., Lefkowitz, R.J., and Milligan, G. (2005). The sustainability of interactions between the orexin-1 receptor and beta-arrestin-2 is defined by a single C-terminal cluster of hydroxy amino acids and modulates the kinetics of ERK MAPK regulation. Biochem. J. *387*: 573–584.

Milligan, G., and Kostenis, E. (2006). Heterotrimeric G-proteins: a short history. Br. J. Pharmacol. *147*: S46–S55.

Mirabeau, O., and Joly, J.-S. (2013). Molecular evolution of peptidergic signaling systems in bilaterians. Proc. Natl. Acad. Sci. U. S. A. *110*: E2028-37.

Miskolzie, M., and Kotovych, G. (2003). The NMR-Derived Conformation of Orexin-A: An Orphan G-Protein Coupled Receptor Agonist Involved in Appetite Regulation and Sleep. J. Biomol. Struct. Dyn. *21*: 201–210.

Molinoff, P.B. (1984). α- and β-Adrenergic Receptor Subtypes. Drugs *28*: 1–15.

Monks, S.A., Karagianis, G., Howlett, G.J., and Norton, R.S. (1996). Solution structure of human neuropeptide Y. J.Biomol.NMR *8*: 379–390.

Moyle, W.R., Xing, Y., Lin, W., Cao, D., Myers, R. V, Kerrigan, J.E., et al. (2004). Model of glycoprotein hormone receptor ligand binding and signaling. J. Biol. Chem. *279*: 44442–44459.

Murphy, P., Moline, M., Mayleben, D., Rosenberg, R., Zammit, G., Pinner, K., et al. (2017). Lemborexant, a dual orexin receptor antagonist (DORA) for the treatment of insomnia disorder: results from a Bayesian, adaptive, randomized,

double-blind, placebo-controlled study. J. Clin. Sleep Med. *13*: 1289–1299.

Nagahara, T., Saitoh, T., Kutsumura, N., Irukayama-Tomobe, Y., Ogawa, Y., Kuroda, D., et al. (2015). Design and Synthesis of Non-Peptide, Selective Orexin Receptor 2 Agonists. J. Med. Chem. *58*: 7931–7937.

Nagarathnam, B., Kalaimathy, S., Balakrishnan, V., and Sowdhamini, R. (2012). Cross-Genome Clustering of Human and *C. elegans* G-Protein Coupled Receptors. Evol. Bioinforma. *8*: EBO.S9405.

Nagase, H., Yamamoto, N., Yata, M., Ohrui, S., Okada, T., Saitoh, T., et al. (2017). Design and Synthesis of Potent and Highly Selective Orexin 1 Receptor Antagonists with a Morphinan Skeleton and Their Pharmacologies. J. Med. Chem. *60*: 1018–1040.

Nathans, J., and Hogness, D.S. (1983). Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. Cell *34*: 807–14.

Nathoo, A.N., Moeller, R.A., Westlund, B.A., and Hart, A.C. (2001). Identification of neuropeptide-like protein gene families in Caenorhabditiselegans and other species. Proc. Natl. Acad. Sci. U. S. A. *98*: 14000–5.

Nelson, J.S., Grande, T.C., and Wilson, M.V.H. (2016). Fishes of the world.

New, D.C., and Wong, J.T. (1998). The evidence for G-protein-coupled receptors and heterotrimeric G proteins in protozoa and ancestral metazoa. Biol. Signals Recept. *7*: 98–108.

Nicole, P., Couvineau, P., Jamin, N., Voisin, T., and Couvineau, A. (2015). Crucial role of the orexin-B C-terminus in the induction of OX1 receptor-mediated apoptosis: analysis by alanine scanning, molecular modelling and site-directed mutagenesis. Br. J. Pharmacol. *172*: 5211–5223.

Noda, K., Saad, Y., Graham, R.M., and Karnik, S.S. (1994). The high affinity state of the beta 2-adrenergic receptor requires unique interaction between conserved and non-conserved extracellular loop cysteines. J. Biol. Chem. *269*: 6743–52.

Nordström, K.J., Fredriksson, R., and Schiöth, H.B. (2008). The amphioxus (Branchiostoma floridae) genome contains a highly diversified set of G protein-coupled receptors. BMC Evol. Biol. *8*: 9.

Nowoshilow, S., Schloissnig, S., Fei, J.-F., Dahl, A., Pang, A.W.C., Pippel, M., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*: 50.

Nygaard, R., Frimurer, T.M., Holst, B., Rosenkilde, M.M., and Schwartz, T.W. (2009). Ligand binding and micro-switches in 7TM receptor structures. Trends Pharmacol. Sci. *30*: 249–259.

Nystén, T., Äystö, L., Laitinen, J., Mehtonen, J., Alhola, K., Leppänen, M., et al. (2019). Environmental drug load can be reduced. SYKE Policy Br. - Näkökulmia

Ympäristöpolitiikkaan *SYKE Polic*:

O'Hara, P.J., Sheppard, P.O., Thógersen, H., Venezia, D., Haldeman, B.A., McGrane, V., et al. (1993). The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. Neuron *11*: 41–52.

Ohlmeyer, M., and Zhou, M.-M. (2010). Integration of Small-Molecule Discovery in Academic Biomedical Research. Mt. Sinai J. Med. A J. Transl. Pers. Med. *77*: 350–357.

Ohno, S. (1970). Evolution by Gene Duplication (Berlin, Heidelberg: Springer Berlin Heidelberg).

Okumura, T., Takeuchi, S., Motomura, W., Yamada, H., Egashira, S., Asahi, S., et al. (2001). Requirement of Intact Disulfide Bonds in Orexin-A-Induced Stimulation of Gastric Acid Secretion That Is Mediated by OX1 Receptor Activation. Biochem. Biophys. Res. Commun. *280*: 976–981.

Opperdoes, F.R., and Lemey, P. (2009). Phylogenetic analysis using protein sequences. In The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing, A.-M. Vandamme, M. Salemi, and P. Lemey, eds. (Cambridge: Cambridge University Press), pp 313–342.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., et al. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. Science *289*: 739–45.

Pándy-Szekeres, G., Munk, C., Tsonkov, T.M., Mordalski, S., Harpsøe, K., Hauser, A.S., et al. (2018). GPCRdb in 2018: adding GPCR structure models and ligands. Nucleic Acids Res. *46*: D440–D446.

Park, K.B., and Weon, H. (2017). Orexin receptors mediate long-term depression of excitatory synaptic transmission in the spinal cord dorsal horn. Neurosci. Lett. *660*: 12–16.

Pasquier, J., Braasch, I., Batzel, P., Cabau, C., Montfort, J., Nguyen, T., et al. (2017). Evolution of gene expression after whole-genome duplication: New insights from the spotted gar genome. J. Exp. Zool. Part B Mol. Dev. Evol. *328*: 709–721.

Pavlos, N.J., and Friedman, P.A. (2017). GPCR Signaling and Trafficking: The Long and Short of It. Trends Endocrinol. Metab. *28*: 213–226.

Pawson, A.J., Faccenda, E., Maudsley, S., Lu, Z.-L., Naor, Z., and Millar, R.P. (2008). Mammalian Type I Gonadotropin-Releasing Hormone Receptors Undergo Slow, Constitutive, Agonist-Independent Internalization. Endocrinology *149*: 1415–1422.

Peer, Y. Van de, Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. *10*: 725–732.

Pellegrini, M., and Mierke, D.F. (1999). Molecular complex of cholecystokinin-8

and N-terminus of the cholecystokinin A receptor by NMR spectroscopy. Biochemistry *38*: 14775–14783.

Peltonen, H.M., Magga, J.M., Bart, G., Turunen, P.M., Antikainen, M.S.H., Kukkonen, J.P., et al. (2009). Involvement of TRPC3 channels in calcium oscillations mediated by OX1 orexin receptors. Biochem. Biophys. Res. Commun. *385*: 408–412.

Perez, F. (2004). Serial Cloner v. 2.6. 0. Softw. Ser.

Peyron, C., Tighe, D.K., Pol, A.N. van den, Lecea, L. de, Heller, H.C., Sutcliffe, J.G., et al. (1998). Neurons containing hypocretin (orexin) project to multiple neuronal systems. J. Neurosci. *18*: 9996–10015.

Piscitelli, C.L., Kean, J., Graaf, C. de, and Deupi, X. (2015). A Molecular Pharmacologist's Guide to G Protein–Coupled Receptor Crystallography. Mol. Pharmacol. *88*: 536 LP – 551.

Plakidou-Dymock, S., Dymock, D., and Hooley, R. (1998). A higher plant seven-transmembrane receptor that influences sensitivity to cytokinins. Curr. Biol. *8*: 315–24.

Postlethwait, J., Amores, A., Cresko, W., Singer, A., and Yan, Y.-L. (2004). Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet. *20*: 481–490.

Price, M.A., Cruzalegui, F.H., and Treisman, R. (1996). The p38 and ERK MAP kinase pathways cooperate to activate Ternary Complex Factors and c-fos transcription in response to UV light. EMBO J. *15*: 6552–6563.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*: 1064–1071.

Putula, J., and Kukkonen, J.P. (2012). Mapping of the binding sites for the OX1 orexin receptor antagonist, SB-334867, using orexin/hypocretin receptor chimaeras. Neurosci. Lett. *506*: 111–115.

Putula, J., Turunen, P.M., Jäntti, M.H., Ekholm, M.E., and Kukkonen, J.P. (2011a). Agonist ligand discrimination by the two orexin receptors depends on the expression system. Neurosci. Lett. *494*: 57–60.

Putula, J., Turunen, P.M., Johansson, L., Näsman, J., Ra, R., Korhonen, L., et al. (2011b). Orexin/hypocretin receptor chimaeras reveal structural features important for orexin peptide distinction. FEBS Lett. *585*: 1368–1374.

PyMOL 1.7.0.0 Schrödinger The PyMOL Molecular Graphics System, Version 1.7.0.0 Schrödinger, LLC.

Qin, K., Dong, C., Wu, G., and Lambert, N.A. (2011). Inactive-state preassembly of G(q)-coupled receptors and G(q) heterotrimers. Nat. Chem. Biol. *7*: 740–747.

Randeva, H.S., Karteris, E., Grammatopoulos, D., and Hillhouse, E.W. (2001). Expression of Orexin-A and Functional Orexin Type 2 Receptors in the Human Adult Adrenals: Implications for Adrenal Function and Energy Homeostasis. J. Clin. Endocrinol. Metab. *86*: 4808–4813.

Ravi, V., and Venkatesh, B. (2008). Rapidly evolving fish genomes and teleost diversity. Curr. Opin. Genet. Dev. *18*: 544–550.

Rhee, S.G., and Bae, Y.S. (1997). Regulation of phosphoinositide-specific phospholipase C isozymes. J. Biol. Chem. *272*: 15045–15048.

Rinne, M., Tanoli, Z.-U.-R., Khan, A., and Xhaard, H. (2019). Cartography of rhodopsin-like G protein-coupled receptors across vertebrate genomes. Sci. Rep. *9*: 7058.

Ritschard, E.A., Fitak, R.R., Oleg, S., and Sönke, J. (2019). Genomic signatures of G-protein-coupled receptor expansions reveal functional transitions in the evolution of cephalopod signal transduction. Proc. R. Soc. B Biol. Sci. *286*: 20182929.

Roecker, A.J., Cox, C.D., and Coleman, P.J. (2016). Orexin Receptor Antagonists: New Therapeutic Agents for the Treatment of Insomnia. J. Med. Chem. *59*: 504–530.

Roecker, A.J., Mercer, S.P., Schreier, J.D., Cox, C.D., Fraley, M.E., Steen, J.T., et al. (2014). Discovery of 5″-Chloro-N-[(5,6-dimethoxypyridin-2-yl)methyl]-2,2′:5′,3″-terpyridine-3′-carboxamide (MK-1064): A Selective Orexin 2 Receptor Antagonist (2-SORA) for the Treatment of Insomnia. ChemMedChem *9*: 311–322.

Rosenbaum, D.M., Rasmussen, S.G.F., and Kobilka, B.K. (2009). The structure and function of G-protein-coupled receptors. Nature *459*: 356.

Rouet-Benzineb, P., Rouyer-Fessard, C., Jarry, A., Avondo, V., Pouzet, C., Yanagisawa, M., et al. (2004). Orexins acting at native OX1 receptor in colon cancer and neuroblastoma cells or at recombinant OX1 receptor suppress cell growth by inducing apoptosis. J. Biol. Chem. *279*: 45875–45886.

Rudolf, J., Dondorp, D., Canon, L., Tieo, S., and Chatzigeorgiou, M. (2019). Automated behavioural analysis reveals the basic behavioural repertoire of the urochordate Ciona intestinalis. Sci. Rep. *9*: 2416.

Ruuskanen, J.O., Xhaard, H., Marjamäki, A., Salaneck, E., Salminen, T., Yan, Y.-L., et al. (2004). Identification of duplicated fourth alpha2-adrenergic receptor subtype by cloning and mapping of five receptor genes in zebrafish. Mol. Biol. Evol. *21*: 14–28.

Ryan, K., Lu, Z., and Meinertzhagen, I.A. (2016). The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. Elife *5*: e16962.

Sacerdot, C., Louis, A., Bon, C., Berthelot, C., and Roest Crollius, H. (2018). Chromosome evolution at the origin of the ancestral vertebrate genome. Genome

Biol. *19*: 166.

Sakurai, T., Amemiya, A., Ishii, M., Matsuzaki, I., Chemelli, R.M., Tanaka, H., et al. (1998). Orexins and Orexin Receptors: A Family of Hypothalamic Neuropeptides and G Protein-Coupled Receptors that Regulate Feeding Behavior. Cell *92*: 573–585.

Salerno, M., Villano, I., Nicolosi, D., Longhitano, L., Loreto, C., Lovino, A., et al. (2019). Modafinil and orexin system: interactions and medico-legal considerations. Front. Biosci. (Landmark Ed. *24*: 564–575.

Sargin, D. (2018). The role of the orexin system in stress response. Neuropharmacology.

Sarkar, A., Kumar, S., and Sundar, D. (2011). The G protein-coupled receptors in the pufferfish Takifugu rubripes. BMC Bioinformatics *12 Suppl 1*: S3.

Satoh, N. (2003). The ascidian tadpole larva: comparative molecular development and genomics. Nat. Rev. Genet. *4*: 285–295.

Schiöth, H.B., and Fredriksson, R. (2005). The GRAFS classification system of G-protein coupled receptors in comparative perspective. Gen. Comp. Endocrinol. *142*: 94–101.

Schneider, E.R., Rada, P., Darby, R.D., Leibowitz, S.F., and Hoebel, B.G. (2007). Orexigenic Peptides and Alcohol Intake: Differential Effects of Orexin, Galanin, and Ghrelin. Alcohol. Clin. Exp. Res. *31*: 1858–1865.

Sea Urchin Genome Sequencing Consortium, S.U.G.S., Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., et al. (2006). The genome of the sea urchin Strongylocentrotus purpuratus. Science *314*: 941–52.

Segner, H., and Baumann, L. (2016). What constitutes a model organism in ecotoxicology? Integr. Environ. Assess. Manag. *12*: 199–200.

Seifert, R., and Wenzel-Seifert, K. (2002). Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors. Naunyn. Schmiedebergs. Arch. Pharmacol. *366*: 381–416.

Semmens, D.C., and Elphick, M.R. (2017). The evolution of neuropeptide signalling: insights from echinoderms. Brief. Funct. Genomics *16*: 288–298.

Semmens, D.C., Mirabeau, O., Moghul, I., Pancholi, M.R., Wurm, Y., and Elphick, M.R. (2016). Transcriptomic identification of starfish neuropeptide precursors yields new insights into neuropeptide evolution. Open Biol. *6*: 150224.

Sharf, R., Sarhan, M., and DiLeone, R.J. (2010). Role of orexin/hypocretin in dependence and addiction. Brain Res. *1314*: 130–138.

Sharman, J.L., Mpamhanga, C.P., Spedding, M., Germain, P., Staels, B., Dacquet, C., et al. (2011). IUPHAR-DB: new receptors and tools for easy searching and visualization of pharmacological data. Nucleic Acids Res. *39*: D534-8.

Shihoya, W., Nishizawa, T., Okuta, A., Tani, K., Dohmae, N., Fujiyoshi, Y., et al. (2016). Activation mechanism of endothelin ETB receptor by endothelin-1. Nature *537*: 363–368.

Shiraishi, A., Okuda, T., Miyasaka, N., Osugi, T., Okuno, Y., Inoue, J., et al. (2019). Repertoires of G protein-coupled receptors for *Ciona* -specific neuropeptides. Proc. Natl. Acad. Sci. 201816640.

Siegel, J.M., and Boehmer, L.N. (2006). Narcolepsy and the hypocretin system—where motion meets emotion. Nat. Clin. Pract. Neurol. *2*: 548–556.

Silve, C., Petrel, C., Leroy, C., Bruel, H., Mallet, E., Rognan, D., et al. (2005). Delineating a Ca2+ binding pocket within the venus flytrap module of the human calcium-sensing receptor. J. Biol. Chem. *280*: 37917–23.

Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., et al. (2015). Hemichordate genomes and deuterostome origins. Nature *527*: 459–65.

Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., et al. (2013). Insights into bilaterian evolution from three spiralian genomes. Nature *493*: 526–31.

Smart, D., Sabido-David, C., Brough, S.J., Jewitt, F., Johns, A., Porter, R.A., et al. (2001). SB-334867-A: the first selective orexin-1 receptor antagonist. Br. J. Pharmacol. *132*: 1179–1182.

Smith, J.J., and Keinath, M.C. (2015). The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. *25*: 1081–1090.

Southan, C. (2004). Has the yo-yo stopped? An assessment of human protein-coding gene number. Proteomics *4*: 1712–1726.

Spring, J. (1997). Vertebrate evolution by interspecific hybridisation – are we polyploid? FEBS Lett. *400*: 2–8.

Sreekumar, K.R., Huang, Y., Pausch, M.H., and Gulukota, K. (2004). Predicting GPCR-G-protein coupling using hidden Markov models. Bioinformatics *20*: 3490–3499.

Steiner, D.F. (1998). The proprotein convertases. Curr. Opin. Chem. Biol. *2*: 31–9.

Steiner, M.A., Gatfield, J., Brisbare-Roch, C., Dietrich, H., Treiber, A., Jenck, F., et al. (2013). Discovery and Characterization of ACT-335827, an Orally Available, Brain Penetrant Orexin Receptor Type 1 Selective Antagonist. ChemMedChem *8*: 898–903.

Steinke, D., Hoegg, S., Brinkmann, H., and Meyer, A. (2006). Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates. BMC Biol. *4*: 16.

Strange, P.G. (2008). Signaling mechanisms of GPCR ligands. Curr. Opin. Drug Discov. Devel. *11*: 196–202.

Strotmann, R., Schröck, K., Böselt, I., Stäubert, C., Russ, A., and Schöneberg, T. (2011). Evolution of GPCR: Change and continuity. Mol. Cell. Endocrinol. *331*: 170–178.

Stump, C.A., Cooke, A.J., Bruno, J., Cabalu, T.D., Gotter, A.L., Harell, C.M., et al. (2016). Discovery of highly potent and selective orexin 1 receptor antagonists (1-SORAs) suitable for in vivo interrogation of orexin 1 receptor pharmacology. Bioorg. Med. Chem. Lett. *26*: 5809–5814.

Summers, C.H., Yaeger, J.D.W., Staton, C.D., Arendt, D.H., and Summers, T.R. (2018). Orexin/hypocretin receptor modulation of anxiolytic and antidepressive responses during social stress and decision-making: Potential for therapy. Brain Res.

Sunahara, R.K., Dessauer, C.W., and Gilman, A.G. (1996). Complexity and Diversity of Mammalian Adenylyl Cyclases. Annu. Rev. Pharmacol. Toxicol. *36*: 461–480.

Suno, R., Kimura, K.T., Nakane, T., Yamashita, K., Wang, J., Fujiwara, T., et al. (2018). Crystal Structures of Human Orexin 2 Receptor Bound to the Subtype-Selective Antagonist EMPA. Structure *26*: 7-19.e5.

Sutherland, E.W. (1971). [Nobel prize in physiology or medicine 1971: the action of hormones outlined]. Lakartidningen *68*: 4991–4995.

Suzuki, N., Hajicek, N., and Kozasa, T. (2009). Regulation and physiological functions of G12/13-mediated signaling pathways. Neurosignals. *17*: 55–70.

Syrovatkina, V., Alegre, K.O., Dey, R., and Huang, X.-Y. (2016). Regulation, Signaling, and Physiological Functions of G-Proteins. J. Mol. Biol. *428*: 3850–3868.

Takai, T., Takaya, T., Nakano, M., Akutsu, H., Nakagawa, A., Aimoto, S., et al. (2006). Orexin-A is composed of a highly conserved C-terminal and a specific, hydrophilic N-terminal region, revealing the structural basis of specific recognition by the orexin-1 receptor. J. Pept. Sci. an Off. Publ. Eur. Pept. Soc. *12*: 443–454.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. *28*: 2731–2739.

Tang, W.J., and Gilman, A.G. (1991). Type-specific regulation of adenylyl cyclase by G protein beta gamma subunits. Science (80-.). *254*: 1500 LP – 1503.

Taylor, J.S., Peer, Y. Van de, Braasch, I., and Meyer, A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *356*: 1661–79.

Thannickal, T.C., Moore, R.Y., Nienhuis, R., Ramanathan, L., Gulyani, S., Aldrich, M., et al. (2000). Reduced Number of Hypocretin Neurons in Human Narcolepsy.

Neuron *27*: 469–474.

Tran, D.-T., Bonaventure, P., Hack, M., Mirzadegan, T., Dvorak, C., Letavic, M., et al. (2011). Chimeric, mutant orexin receptors show key interactions between orexin receptors, peptides and antagonists. Eur. J. Pharmacol. *667*: 120–128.

Tsuneki, H., Murata, S., Anzawa, Y., Soeda, Y., Tokai, E., Wada, T., et al. (2008). Age-related insulin resistance in hypothalamus and peripheral tissues of orexin knockout mice. Diabetologia *51*: 657–667.

Turku, A., Borrel, A., Leino, T.O., Karhu, L., Kukkonen, J.P., and Xhaard, H. (2016). Pharmacophore Model To Discover OX $_1$ and OX $_2$ Orexin Receptor Ligands. J. Med. Chem. *59*: 8263–8275.

Turku, A., Leino, T.O., Karhu, L., Yli-Kauhaluoma, J., Kukkonen, J.P., Wallén, E.A.A., et al. (2019). Structure–Activity Relationships of 1-Benzoylazulenes at the OX1 and OX2 Orexin Receptors. ChemMedChem *14*: 965–981.

Turunen, P.M., Jäntti, M.H., and Kukkonen, J.P. (2012). OX1 orexin/hypocretin receptor signaling through arachidonic acid and endocannabinoid release. Mol. Pharmacol. *82*: 156–167.

Urbańska, A., Sokołowska, P., Woldan-Tambor, A., Biegańska, K., Brix, B., Jöhren, O., et al. (2012). Orexins/hypocretins acting at G i protein-coupled OX 2 receptors inhibit cyclic AMP synthesis in the primary neuronal cultures. J. Mol. Neurosci. *46*: 10–17.

Veenstra, J.A. (2010). Neurohormones and neuropeptides encoded by the genome of Lottia gigantea, with reference to other mollusks and insects. Gen. Comp. Endocrinol. *167*: 86–103.

Veenstra, J.A. (2011). Neuropeptide evolution: Neurohormones and neuropeptides predicted from the genomes of Capitella teleta and Helobdella robusta. Gen. Comp. Endocrinol. *171*: 160–175.

Venkatakrishnan, A.J., Deupi, X., Lebon, G., Heydenreich, F.M., Flock, T., Miljus, T., et al. (2016). Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. Nature *536*: 484–487.

Venkatesh, B., Lee, A.P., Ravi, V., Maurya, A.K., Lian, M.M., Swann, J.B., et al. (2014). Elephant shark genome provides unique insights into gnathostome evolution. Nature *505*: 174–179.

Vernier, P., Cardinaud, B., Valdenaire, O., Philippe, H., and Vincent, J.D. (1995). An evolutionary view of drug-receptor interaction: the bioamine receptor family. Trends Pharmacol. Sci. *16*: 375–81.

Wang, C., Wu, H., Katritch, V., Han, G.W., Huang, X.-P., Liu, W., et al. (2013). Structure of the human smoothened receptor bound to an antitumour agent. Nature *497*: 338–43.

Wang, P., Wang, M., Zhang, L., Zhong, S., Jiang, W., Wang, Z., et al. Functional characterization of an orexin neuropeptide in amphioxus reveals an ancient origin of orexin/orexin receptor system in chordate. Sci. CHINA Life Sci.

Ward, R.J., Pediani, J.D., and Milligan, G. (2011). Ligand-induced internalization of the orexin OX(1) and cannabinoid CB(1) receptors assessed via N-terminal SNAP and CLIP-tagging. Br. J. Pharmacol. *162*: 1439–1452.

Webb, B., and Sali, A. (2014). Protein structure modeling with MODELLER. Methods Mol. Biol. *1137*: 1–15.

Wegener, C., and Gorbashov, A. (2008). Molecular evolution of neuropeptides in the genus Drosophila. Genome Biol. *9*: R131.

Wellcome Sanger Institute (2019). Genetic code of 66,000 UK species to be sequenced.

Wheatley, M., Wootten, D., Conner, M.T., Simms, J., Kendrick, R., Logan, R.T., et al. (2012). Lifting the lid on GPCRs: the role of extracellular loops. Br. J. Pharmacol. *165*: 1688–1703.

White, J.F., Noinaj, N., Shibata, Y., Love, J., Kloss, B., Xu, F., et al. (2012). Structure of the agonist-bound neurotensin receptor. Nature *490*: 508–513.

Willie, J.T., Chemelli, R.M., Sinton, C.M., Tokita, S., Williams, S.C., Kisanuki, Y.Y., et al. (2003). Distinct Narcolepsy Syndromes in Orexin Receptor-2 and Orexin Null Mice: Molecular Genetic Dissection of Non-REM and REM Sleep Regulatory Processes. Neuron *38*: 715–730.

Willie, J.T., Renthal, W., Chemelli, R.M., Miller, M.S., Scammell, T.E., Yanagisawa, M., et al. (2005). Modafinil more effectively induces wakefulness in orexin-null mice than in wild-type littermates. Neuroscience *130*: 983–995.

Winrow, C.J., Gotter, A.L., Cox, C.D., Doran, S.M., Tannenbaum, P.L., Breslin, M.J., et al. (2011). Promotion of Sleep by Suvorexant—A Novel Dual Orexin Receptor Antagonist. J. Neurogenet. *25*: 52–61.

Winrow, C.J., Gotter, A.L., Cox, C.D., Tannenbaum, P.L., Garson, S.L., Doran, S.M., et al. (2012). Pharmacological characterization of MK-6096 – A dual orexin receptor antagonist for insomnia. Neuropharmacology *62*: 978–987.

Wisor, J.P., Nishino, S., Sora, I., Uhl, G.H., Mignot, E., and Edgar, D.M. (2001). Dopaminergic Role in Stimulant-Induced Wakefulness. J. Neurosci. *21*: 1787 LP – 1794.

Woldan-Tambor, A., Biegańska, K., Wiktorowska-Owczarek, A., and Zawilska, J.B. (2011). Activation of orexin/hypocretin type 1 receptors stimulates cAMP synthesis in primary cultures of rat astrocytes. Pharmacol. Reports *63*: 717–723.

Wolf, S., and Grünewald, S. (2015). Sequence, Structure and Ligand Binding Evolution of Rhodopsin-Like G Protein-Coupled Receptors: A Crystal Structure-

Based Phylogenetic Analysis. PLoS One *10*: e0123533.

Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G.W., Vardy, E., et al. (2012). Structure of the human kappa-opioid receptor in complex with JDTic. Nature *485*: 327–332.

Xhaard, H., Rantanen, V.-V., Nyrönen, T., and Johnson, M.S. (2006). Molecular evolution of adrenoceptors and dopamine receptors: implications for the binding of catecholamines. J. Med. Chem. *49*: 1706–19.

Xia, J.X., Fan, S.Y., Yan, J., Chen, F., Li, Y., Yu, Z.P., et al. (2009). Orexin A-induced extracellular calcium influx in prefrontal cortex neurons involves L-type calcium channels. J. Physiol. Biochem. *65*: 125–136.

Xu, M., and Volkoff, H. (2007). Molecular characterization of prepro-orexin in Atlantic cod (Gadus morhua): Cloning, localization, developmental profile and role in food intake regulation. Mol. Cell. Endocrinol. *271*: 28–37.

Yamamoto, K., Mirabeau, O., Bureau, C., Blin, M., Michon-Coudouel, S., Demarque, M., et al. (2013). Evolution of Dopamine Receptor Genes of the D1 Class in Vertebrates. Mol. Biol. Evol. *30*: 833–843.

Yamanaka, N., Yamamoto, S., Žitňan, D., Watanabe, K., Kawada, T., Satake, H., et al. (2008). Neuropeptide receptor transcriptome reveals unidentified neuroendocrine pathways. PLoS One *3*:.

Yanagisawa, M. (2010). Small-molecule agonists for type-2 orexin receptor. Patent US201001500840A1.

Yegorov, S., and Good, S. (2012). Using paleogenomics to study the evolution of gene families: origin and duplication history of the relaxin family hormones and their receptors. PLoS One *7*: e32923.

Yin, J., Babaoglu, K., Brautigam, C.A., Clark, L., Shao, Z., Scheuermann, T.H., et al. (2016a). Structure and ligand-binding mechanism of the human OX1 and OX2 orexin receptors. Nat.Struct.Mol.Biol. *23*: 293–299.

Yin, J., Babaoglu, K., Brautigam, C.A.C.A., Clark, L., Shao, Z., Scheuermann, T.H.T.H., et al. (2016b). Structure and ligand-binding mechanism of the human OX1 and OX2 orexin receptors. Nat.Struct.Mol.Biol. *23*: 293–299.

Yin, J., Mobarec, J.C., Kolb, P., and Rosenbaum, D.M. (2015). Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. Nature *519*: 247–250.

Yoshida, Y., Fujiki, N., Nakajima, T., Ripley, B., Matsumura, H., Yoneda, H., et al. (2001). Fluctuation of extracellular hypocretin-1 (orexin A) levels in the rat in relation to the light–dark cycle and sleep–wake activities. Eur. J. Neurosci. *14*: 1075–1081.

Yoshida, Y., Terauchi, T., Naoe, Y., Kazuta, Y., Ozaki, F., Beuckmann, C.T., et al.

(2014). Design, synthesis, and structure–activity relationships of a series of novel N-aryl-2-phenylcyclopropanecarboxamide that are potent and orally active orexin receptor antagonists. Bioorg. Med. Chem. *22*: 6071–6088.

Yun, S., Furlong, M., Sim, M., Cho, M., Park, S., Cho, E.B., et al. (2015). Prevertebrate Local Gene Duplication Facilitated Expansion of the Neuropeptide GPCR Superfamily. Mol. Biol. Evol. *32*: 2803–17.

Zeng, F.Y., Soldner, A., Schöneberg, T., and Wess, J. (1999). Conserved extracellular cysteine pair in the M3 muscarinic acetylcholine receptor is essential for proper receptor cell surface localization but not for G protein coupling. J. Neurochem. *72*: 2404–14.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*: D754–D761.

Zhang, S., Blache, D., Vercoe, P.E., Adam, C.L., Blackberry, M.A., Findlay, P.A., et al. (2005). Expression of orexin receptors in the brain and peripheral tissues of the male sheep. Regul. Pept. *124*: 81–87.

Recent Publications in this Series

46/2019 Suvi Koskinen
Near-Occlusive Atherosclerotic Carotid Artery Disease: Study with Computed Tomography Angiography
47/2019 Flavia Fontana
Biohybrid Cloaked Nanovaccines for Cancer Immunotherapy
48/2019 Marie Mennesson
Kainate Receptor Auxiliary Subunits Neto1 and Neto2 in Anxiety and Fear-Related Behaviors
49/2019 Zehua Liu
Porous Silicon-Based On-Demand Nanohybrids for Biomedical Applications
50/2019 Veer Singh Marwah
Strategies to Improve Standardization and Robustness of Toxicogenomics Data Analysis
51/2019 Iryna Hlushchenko
Actin Regulation in Dendritic Spines: From Synaptic Plasticity to Animal Behavior and Human Neurodevelopmental Disorders
52/2019 Heini Liimatta
Effectiveness of Preventive Home Visits among Community-Dwelling Older People
53/2019 Helena Karppinen
Older People´s Views Related to Their End of Life: Will-to-Live, Wellbeing and Functioning
54/2019 Jenni Laitila
Elucidating Nebulin Expression and Function in Health and Disease
55/2019 Katarzyna Ciuba
Regulation of Contractile Actin Structures in Non-Muscle Cells
56/2019 Sami Blom
Spatial Characterisation of Prostate Cancer by Multiplex Immunohistochemistry and Quantitative Image Analysis
57/2019 Outi Lyytinen
Molecular Details of the Double-Stranded RNA Virus Replication and Assembly
58/2019 Markus Räsänen
Vascular Endothelial Growth Factor-B and the Bmx Tyrosine Kinase in Cardiac Hypertrophy and Revascularization
59/2019 Vuokko Nummi
Insights into Clinical and Laboratory Phenotypes of Von Willebrand Disease
60/2019 Shah Hasan
Challenges of Hyper-Prolificacy in the Pig: Colostrum and Gut Microbiota
61/2019 Sanna Matilainen
Pathomechanisms of Leigh Syndrome: Defects of Post-Transcriptional and Post-Translational Regulation of Mitochondrial Metabolism
62/2019 Kirsi Santti
Desmoid Tumor: Oncological Management and Prognostic Biomarkers
63/2019 Hesham E. Abdolhfid Mohamed
Evaluation of Prognostic Markers for Oropharyngeal Carcinoma Using Tissue Microarray
64/2019 Johanna Uhari-Väänänen
Contributions of μ- and κ-Opioidergic Systems to Ethanol Intake and Addiction
65/2019 Susanna Rapo-Pylkkö
Chronic Pain and Neuropathic Pain among Community-dwelling Older Adults in Primary Health Care Settings
66/2019 Helka Göös
Human Transcription Factor Protein-protein Interactions in Health and Disease

Helsinki 2019 **ISSN 2342-3161** **ISBN 978-951-51-5468-2**