1 INTRODUCTION

1.1 The Thyroid Stimulating Hormone Receptor

In this section the classification and medical importance of the Thyroid Stimulating Hormone Receptor (TSHR) will be described.

The TSHR belongs to the family of G-protein coupled receptors (GPCRs) mediating most of their intracellular actions through G-proteins. G-proteins transmit the signal to effector proteins, such as enzymes and ion channels, resulting in rapid changes in the concentration of intracellular signalling molecules, cAMP, cGMP, inisitol phosphates, diacylglycerol, arachidonic acid and cytosolic ions (Kristiansen K 2004).

The GPCRs are classified into several families (Kolakowski LF 1994, Fredriksson R 2003), where more than 90% of all GPCRs are grouped into the rhodopsin/adrenergic–like receptor family A. Members of the superfamily of GPCRs share a common molecular architecture. Structural data on GPCRs based on biochemical, immunological and biophysical approaches have validated a consensus topology of GPCRs with an extracellular N-terminus, a cytoplasmatic C-terminus and seven transmembrane helices (TMHs) connected by intra- and extracellular loops (ICLs, ECLs). The X-ray structures of bovine rhodopsin (Palczewski K 2000, Li J 2004, Okada T 2004) (pdb entry codes: 1F88, 1HZX, 1L9H) are the only crystal structures of any GPCRs.

This superfamily includes receptors for diverse endogenous ligands including amines, peptides, amino acids, nucleosides, Ca^{2+} ions and sensory receptors for various exogenous ligands such as odorants, pheromones, and photons of light (Schwartz TW 2006). Dysfunction of the GPCRs causes human diseases. Therefore many GPCRs are targets for pharmaceuticals and drugs (Schöneberg T 2004). It has been estimated that GPCRs represent over 40% of current drug targets (Flower DR 1999).

The thyroid stimulating hormone (TSH, thyrotropin) receptor (TSHR) together with the lutropin (LH), choriogonadotropin (CG) and follitropin (FSH) receptors (LHCGR and FSHR) belongs to the subfamily of glycoprotein-hormone receptors (GPHRs) within the rhodopsin/adrenergic receptor family in the GPCR superfamily (Kosugi S 1996 (a), Themmen APN 2000, Ascoli M 2002). TSH and the TSH receptor are key proteins in the control of thyroid function. TSH synthesis in the anterior pituitary gland is inhibited by thyroid hormone in a classical endocrine negative-feedback loop. TSH binds to its receptor on thyroid cells and leads to the stimulation of secondary messenger pathways involving predominantly cAMP

(Laugwitz KL 1996). Inositol 1,4,5-triphosphate (IP_3) and diaglycerol (DAG) pathways are also activated at 10-fold higher concentrations of TSH (van Sande J 2006). The physiological roles of TSH include stimulation of differentiated thyroid functions, such as iodine uptake and organification, production and release of iodothyronines from the gland, and promotion of thyroid growth. TSH also acts as a factor protecting thyroid cells from apoptosis and plays a critical role in ontogeny (Szkudlinski MW 2002).

The TSHR gene on chromosome 14q3 encodes for a 764 amino acid protein, which comprises a signal peptide of 21 amino acids; a large, glycosylated ectodomain (ECD) of 394 residues encoded by 9 exons; and 349 residues encoded by the tenth and largest exon, which constitute the 7 TMHs and cytoplasmatic tail (Szkudlinski MW 2002).

The TSHR is activated by TSH, Thyrostimulin (Nakabayashi K 2002), constitutively activating mutations (CAMs) (Rodien P 2003), mutations causing promiscuous hormone binding (Costagliola S 2005), antibodies (Davies TF 2005), tryptic action (van Sande J 1996, Chen CR 2003 (a)), small ligands (Jäschke H 2006 (a)), and deletions of epitopes in the extracellular domain (Zhang ML 1995) or the serpentine domain (SD) (Wonerow P 1998). The activation by mild tryptic action, antibodies or mutational deletions are specific events for the TSHR compared to LHCGR and FSHR. Moreover, the TSHR is characterized by a high level of basal activity compared to the homologous FSHR and LHCGR (Simoni M 1997, Dufau ML 1998).

The thyroid stimulating hormone is a 28- to 30-kDa glycoprotein synthesized and secreted from thyrotrophs of the anterior pituitary gland. It is a member of the glycoprotein hormone family that includes FSH, LH and CG. Glycoprotein hormones are among the largest endocrine ligands known to date and they are endogenously acting ligands for their corresponding receptors. They are heterodimeric cysteine-knot glycoproteins consisting of a common α-subunit and a unique β-subunit, which confers biological specificity to each hormone (Hearn MT 2002).

Dysfunction of the TSHR causes several diseases including hyper- and hypothyroidism resulting from germline and somatic mutations, TSHR autoimmunity and hormone binding promiscuity (Schöneberg T 2004, Costagliola S 2005). Hyperthyroidism is a frequent disease, but until now pharmaceutical therapies do not target the TSHR directly. Therefore, the delineation of the molecular activation mechanism of the TSHR and the discovery of sites for pharmacological intervention such as by low molecular weight ligands (antagonists or inverse

agonists) could enable new therapeutic strategies directly affecting the TSHR as origin of such diseases.

1.2 Structural architecture of the TSHR and previously identified inter-/ intramolecular functionalities of extracellular receptor components

This section will summarize knowledge concerning the structural architecture of the TSHR. Extracellular features will be described in detail and in relation to previously identified functional specificities.

The amino acid sequences of the GPHR family are highly conserved (Vassart G 2004 (a)). The common membrane topology of the TSHR and homologous FSHR and LHCGR are shared with the members of the superfamily A of GPCRs. They contain an extracellular N-terminus, a cytoplasmatic C-terminus and seven TMHs connected by intra- and extracellular loops. Only the hinge region connecting the leucine-rich-repeat domain (LRRD) with the SD, and the ECLs of the transmembrane domain are strongly diverse in their amino acid sequence between the three GPHR subclasses (Figure 1.1).

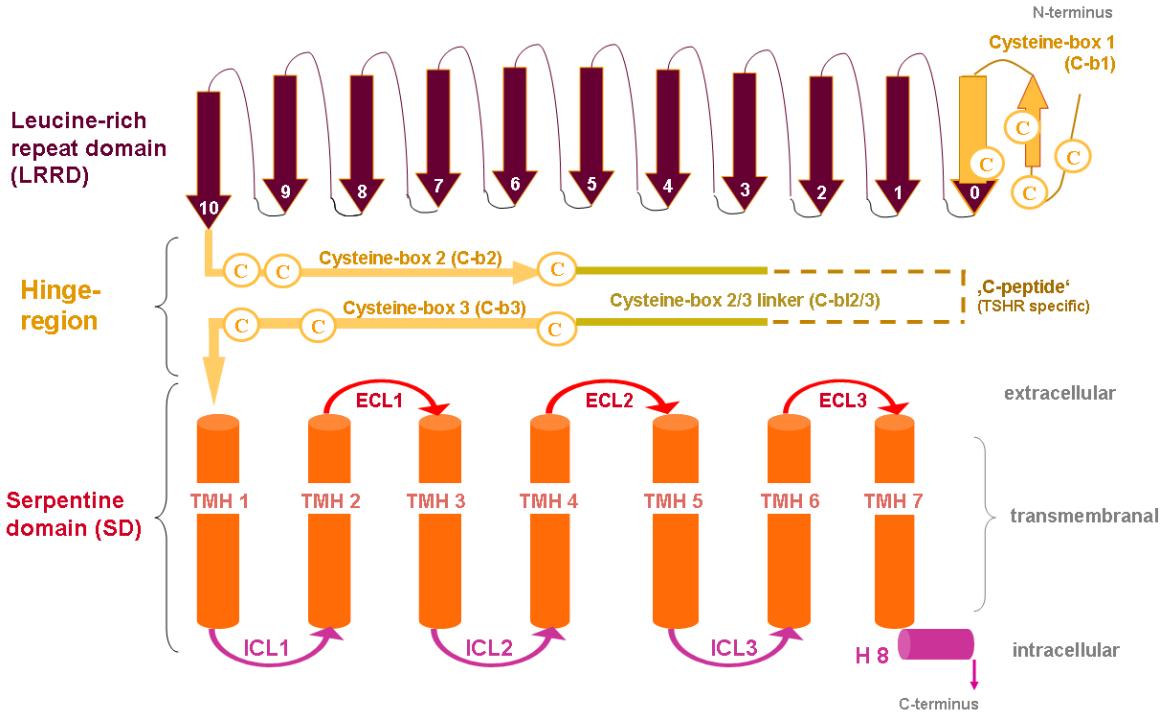

Figure 1.1: ***General scheme of GPHRs***

The common structural characteristics of the large N-terminal ECD (M1-D410) of GPHRs can be subdivided into: i. The C-b1, ii. the LRR domain, iii. the hinge region, which can be subdivided into vi. the C-b2, v. the C-bl2/3, and vi. the C-b3. The SD of GPHRs consists of three ECLs 1-3, the TMHs 1-7, and three ICLs 1-3.

1.2.1 Extracellular structure of the TSHR

1.2.1.1 N-terminal structure

Based on the FSHR LRRD crystal structure (Fan QR 2005 (a)) and specificities in the amino acid sequence the N-terminal ECD of the TSHR (M1–D410) can be subdivided into (all numbers are hTSHR specific): i. the extreme N-terminal tail (including the signal peptide), ii. the cysteine-box 1 (C-b1; C24-C41), iii. the eleven leucine rich repeats forming the leucine rich repeat domain (LRRD; P37 - Y279) and iv. the hinge region (P280 – D410) that can be subdivided further into vi-a. the cysteine-box 2 (C-b2; P280-C301) vi-b. the cysteine-box 2/3 linker region (C-bl 2/3; N302-I389) and vi-c. the cysteine-box 3 (C-b3; C390-D410) located close to TMH1.

1.2.1.1.1 Cysteine-box 1

Recently the crystal structure of the FSHR LRRD in complex with the hormone FSH was solved (Fan QR 2005 (a)). The bordering amino acids in the crystal structure are C18 (N-terminal) and Y250 (C-terminal) (in TSHR C24 and W258). This crystal structure provides clear evidence for the general GPHR LRRD topology and folding of this domain and also offers structural insights in the first cysteine-box (C-b1) out of three cysteine-boxes of the N-terminal domain. C-b1 is an integral part of the LRRD (Figure 1.2).

Figure 1.2*: Crystal structure of the FSHR leucine rich repeat domain*
Red marked is the cysteine-box 1 as an integral part of the LRRD characterized by an anti-parallel β-strand and by 4 cysteines interacting via disulfide bridges stabilizing the LRRD fold at the N-terminus (Fan QR 2005 (a))

In cysteine-box 1 four cysteines are linked pairwise via disulfide bridges and stabilize the N-termius of the LRRD. These four disulfide-bridged cysteines are the 'N-cap' motif (Figure 1.3) that is typical for many LRRDs (Kajava AV 2002).

1.2.1.1.2 Leucine-rich repeat domain

For GPHRs it has been demonstrated that the major binding region for the large hormones LH/CG, TSH and FSH is located in the extracellular leucine rich repeat domain (Braun T

1992). LRRD proteins have been identified in viruses, bacteria, archae and eukaryotes (Enkbayar P 2003). Many LRRDs are involved in protein-protein and protein-ligand interactions, including plant immune response and mammalian innate immune response. Most repeats are 20-30 amino acids long and contain a conserved 11 residue segment with the consensus sequence LxxLxLxx(N/C)xL (x – L, V, I, or F), and the repeat number ranges from 2 to 52. The LRRDs have been divided into seven classes characterized by different lengths and consensus sequences of the variable segments of repeats (Enkbayar P 2003).

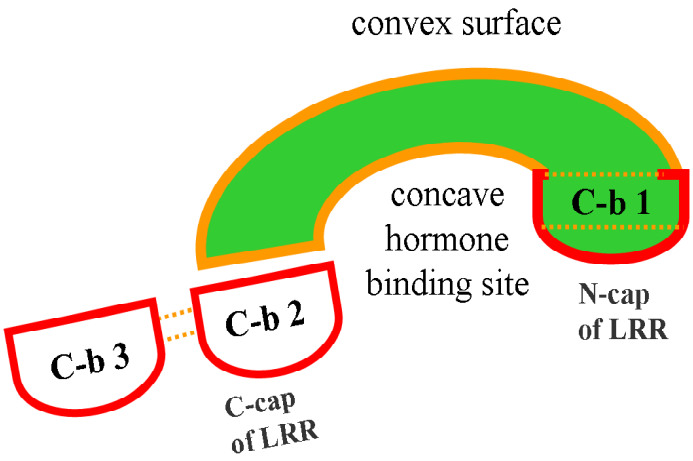

Figure 1.3: ***Scheme of the leucine rich repeat domain***
The two cysteine-boxes C-b1 (N-cap) and C-b2 (C-cap) which flank and stabilize the fold of the LRRD and C-b3 are red boxed. Additionally, C-b1 is an integral part of the LRRD structure (orange bordered) and is characterized by 4 cysteines interacting by disulfide bridges (dashed orange lines). Cysteines of C-b2 are most likely disulfide linked to C-b3 at the C-terminal part of the ECD near TMH1.

The crystal structures of proteins containing LRRDs show that the domain folds into a curved shape with a parallel β-sheet on the concave face and with various secondary structures, including α-helix, 3-10 helix and pII helix on the convex face. The LRRD are arranged in horseshoe- or J-shaped molecules. Exposed side chains on the concave face provide selective interaction patterns for hormone-receptor binding (Braun T 1991).

The crystal structure of the FSHR LRRD (Fan QR 2005 (a)) reveals that LRRs 1-7 lie nearly flat, whereas 7-10 have a larger signature arch-like curvature since the convex side of the FSHR LRRD contains no helices (Figure 1.4). This results in a 'scythe-blade' shape of the GPHR LRRDs that was previously suggested in a homology model by Kleinau et al. (Kleinau G 2004) based on the sequence similarity to the hNogo-receptor ectodomain structure (pdb entry code 1OZN) before the crystal structure of the FSHR LRRD was published.
The N-terminal LRRD also occurs in the closely related leucine rich repeat G-protein coupled receptors (LGRs) 7-8 (Sudo S 2003 (a)). In these receptors they are also essential for binding of the peptide hormones (relaxins 1 to 3 and insulin-like 3/relaxin-like factor). Several orphan homologous receptors of this LGR subgroup (LGR4-6) that contain LRRDs in their N-

terminus also share special finger print amino acids in cysteine-boxes 2 and 3 and within the transmembrane domain.

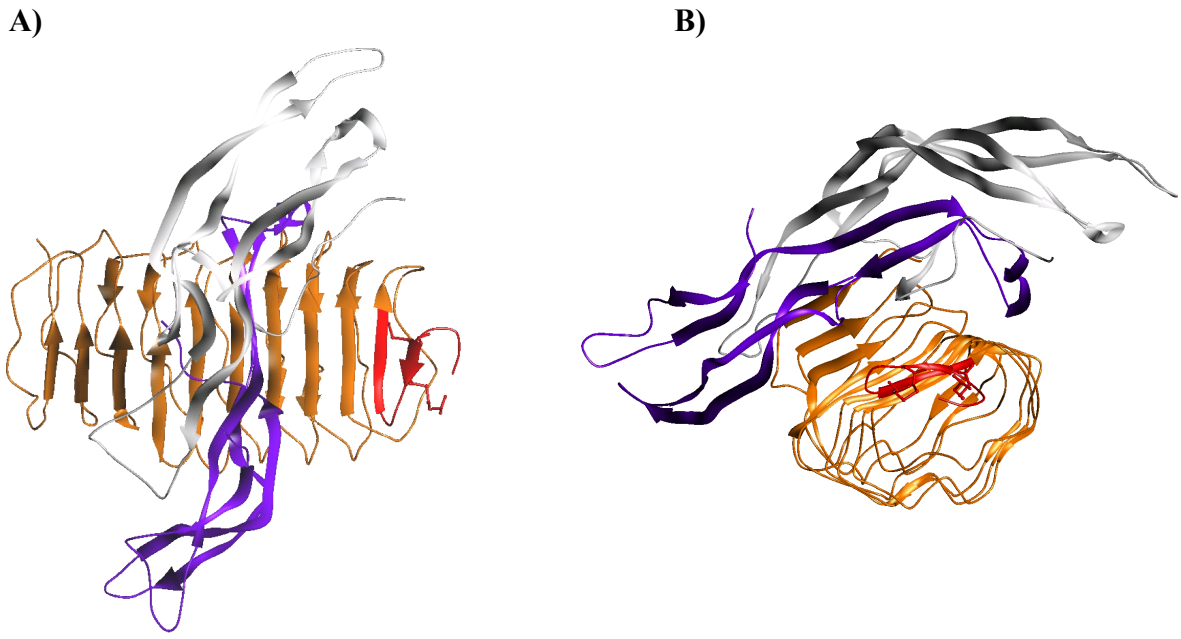

Figure 1.4: ***The crystalized 2.5Å FSHR LRR domain - FSH complex***
FSH in complex with the LRR domain of FSHR: purple - αFSH, white - βFSH, orange – LRRD, red – cysteine-box 1. A) view with hormone in front, B) hormone on top. The hormone is in contact with the β strands in the concave hormone-binding site, mainly via electrostatic interactions. Cysteine-box 1 is an integral part of the LRRD (Fan QR 2005 (a)).

1.2.1.1.3 Hinge region

The region between the LRRD (see section *1.2.1.1.2*) and the serpentine domain is named the *hinge region* (~130 amino acids) (Figure 1.1). The hinge region consists of N- and C-terminally located cysteine-boxes and the sequences of the two cysteine-boxes are connected via a sequence region that is called the cysteine-box 2/3 linker region.

1.2.1.1.3.1 Cysteine-box 2

The cysteine-box 2 (C-b2) represents the 'C-cap' motif of the LRRD (Figure 1.3) that is characterized by three cysteines and flanks the LRRD at the C-terminus. C-b2 is most probably involved in stabilizing the folding of the LRRD.

The conserved S281 of the TSHR is located at the N-terminal end of C-b2 (Figure 1.5). The importance of this position arises because several different pathogenic mutations were identified for it (Duprez L 1997, Kopp P 1997).

In vitro studies have shown that most mutations of this serine lead to constitutive activation (Nakabayashi K 2000, Ho SC 2001), which demonstrates the importance of this serine for the stabilization of the basal receptor activity state.

Substitutions of the corresponding S277 in the LHCGR by all other 19 amino acids increase basal activity by varying degrees. For the epitope Y279-S281 a turn conformation is assumed (Nakabayashi K 2003).

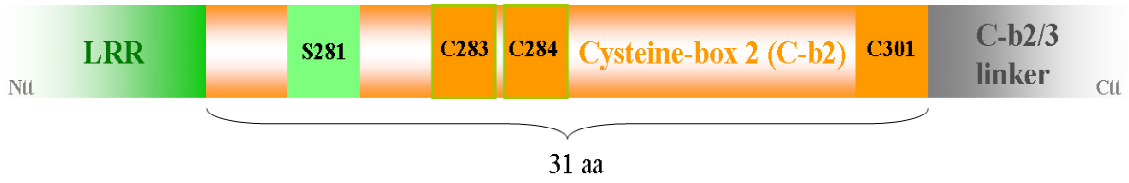

Figure 1.5: ***hTSHR sequence region of Cysteine-box 2***
The cysteine-box 2 (C-b2) consists of 31 amino acids and is an integral part of the hinge-region. C-b2 follows the LRR domain back-to-back. Pathogenic constitutively activating mutations at S281 (light green box) underline the importance of this region for receptor function. Mutations at C283 and C284 also lead to constitutive activity (light green boxed) and these cysteines are additionally essential for receptor protein folding (Ho SC 2001).

This is confirmed by substitution of proline residue P276 (TSHR: P280) adjacent to S277 (TSHR: S281) of the LHCGR with a glycine, which exerts less structural strain and leads to constitutive receptor activation. Mutation of the TSHR residues C283 and C284 to serine leads to a 2-fold increase of basal TSHR activity (Ho SC 2001). In addition, adjacent amino acids are reported as signalling sensitive in the LHCGR (TSHR D276, L277, S278, P280, H282, A285, F286) (Zeng H 2001), whereas P276A,G mutations in the LHCGR (TSHR P280) are also constitutively active *in vitro*.

Disulfide bonds are involved in the quaternary structure of the TSHR (Rapoport B 1998). It has been established that C-b2 potentially interacts with C-b3 close to TMH1 via disulfide bonds (Nagayama Y 1991, Tanaka K 1998), and that mutations disrupting the disulfide-bridges lead to constitutive activation (Ho SC 2003). Currently, there are no direct data to indicate which of the single cysteines form pairs, however, a certain amount of indirect evidence permits the global assignment of disulfide bridges (Zhang R 1996, Bozon V 2002). Mutations of C283, C284 (C-b2) and C398, C408 (C-b3) have dramatic effects on the TSHR structure and TSH binding.

The hypothesis that C301 (C-b2) and C390 (C-b3) are paired is supported by the fact that mutations of either C301 or C390 produce the identical effect of reduced TSH binding affinity (Nagayama Y 1992). Moreover, the LHCG receptor of *Callythrix jacchus* is missing exon 10

which encodes that particular cysteine corresponding to C301 of the TSHR (Gromoll J 2003). Additionally, this receptor simultaneously misses the cysteine corresponding to C390 of the TSHR. Taken together, the two neighbouring cysteines C283 and C284 in C-b2 are very likely to be paired either to C398 and C408 in C-b3 or in reversed order to C408 and C398, whereas C301 (in C-b2) is very likely to be paired with C390 (in C-b3).

One additional structural-functional feature of C-b2 is related to a cluster of positively charged amino acids. Mutagenesis studies revealed tryptic clipping within C-b2 in a cluster of positively charged amino acids (K287-R293) close to S281 that also leads to receptor activation (Chen CR 2003 (b)).

1.2.1.1.3.2 Cysteine-box 2/3 linker

The cysteine-box 2/3 linker region is located between the last and first cysteine of C-b2 and C-b3 (Figure 1.6), respectively, and it is diverse in the length and composition of its amino acid sequence between GPHR subtypes.

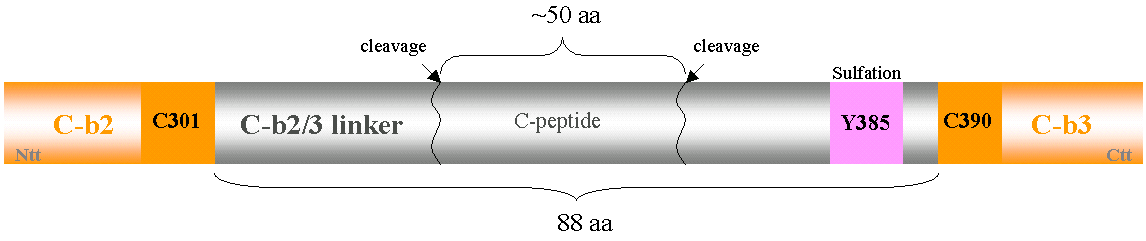

Figure 1.6: ***Cysteine-box 2/3 linker region of hTSHR***
The cysteine-box 2/3 linker region comprises 88 amino acids and is N-terminally and C-terminally bordered by C-b2 and C-b3. Two cleavage sites define the length of the C-peptide (~50 amino acids). Sulfation (negative charge) at amino acid Y385 (lilac) influences hormone binding (Costagliola S 2002).

Three functional-structural features define the C-b2/3 linker region: 1.) A ~50 amino acid long insertion (C-peptide) is significant in the TSHR compared to the FSHR and the LHCGR (Figure 1.6). The functional role *in vivo* is not clarified yet (Vassart G 2004 (b)).

2.) It was assumed, that a single-chain TSHR monomer is posttranslationally cleaved at the cell surface by a metalloprotease (Couet J 1996 (a)). The resulting protomers, termed TSHR α (or subunit A) for the N-terminal region and TSHR β (or subunit B) for the C-terminal region, are held together by disulfide bonds between C-b2 and C-b3 (section *1.2.1.1.3.1*) (Figure 1.7). The cleavage-process of the TSHR into subunits A and B is related to two defined regions within the C-b2/3 linker region epitopes at the N- and C-terminal parts of the TSHR specific

amino acid insertion (N316-Q367) (Chazenbalk GD 1997). After cleavage has occured, the protein disulfide isomerase (Couet J 1996 (b)) sheds subunit A from subunit B by disruption of the disulfide bridges between C-b2 and C-b3, and thus causes a certain amount of subunit A to be released. However, solid evidence for a functional difference between single-chain and two-chain receptors does not currently exist. Moreover, cleavage into subunits does not appear to be required for hormone-induced receptor activation (Chazenbalk GD 2001), and recently, it was confirmed that the TSHR is present *in vivo* as an uncleaved receptor (Chen CR 2006).

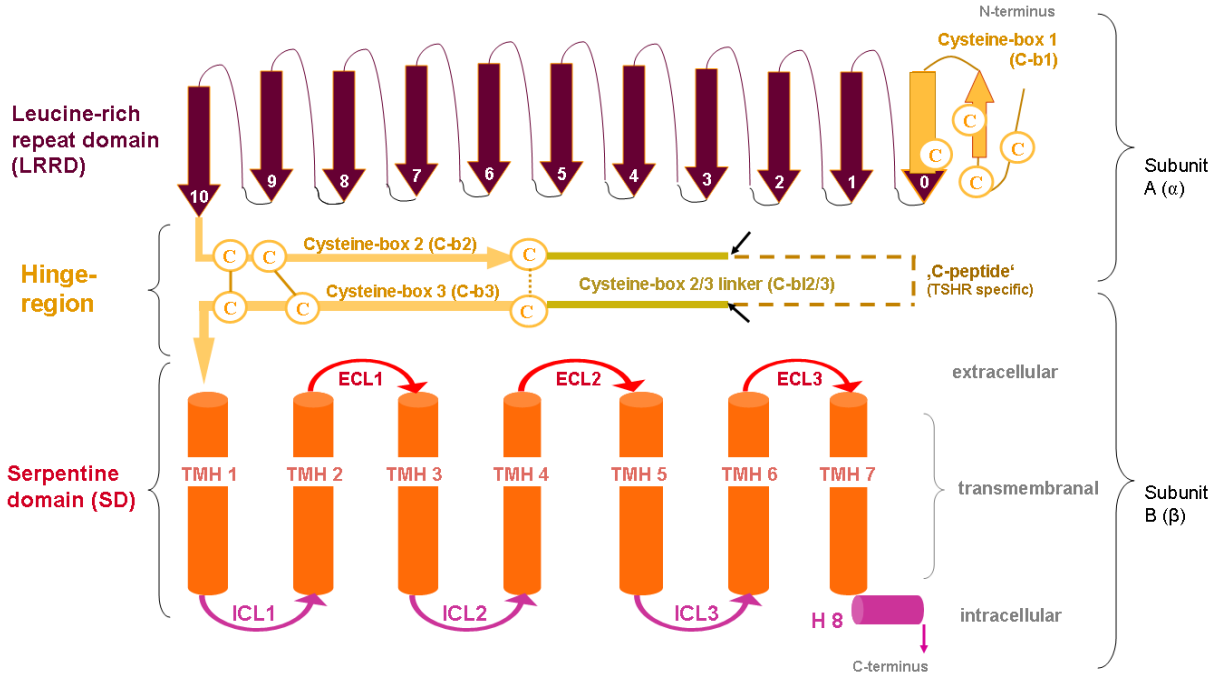

Figure 1.7: ***Cysteine-boxes 2 and 3 are linked via disulfide bridges***
The cysteine-boxes 2 and 3 in the hinge region are most probably linked via disulfide bridges. Disruption of these interactions by a disulfide isomerase ('shedding') leads to constitutive activation of the receptor. Sites of cleavage are located insight the cysteine-box 2,3 linker region (arrows).

3.) Strong indication for the fact that a short conserved epitope of five amino acids is important for natural ligand binding in all GPHRs (D382-Y385) was provided by studies of Kosugi et al. (Kosugi S 1991) and Costagliola et al (Costagliola S 2002). Sulfation of specific tyrosines in the N-terminal domain of CCR5 was shown to contribute to the binding of its natural ligands (Blanpain C 1995). Similarly, sulfated tyrosines are involved in the formation of the docking site for the human C5a receptor (Farzahn M 2001).

Taken together, the functional and structural features of the C-b2/3 linker region are of high importance for receptor functions like hormone binding and signal transmission, but the general context and detailed mechanisms are not well understood yet.

1.2.1.1.3.3 Cysteine-box 3

It has been shown by mutation analysis (Huang JD 1995, Stavrou SS 1998) that the amino acids E409 and D410 at the extreme C-terminal hinge region close to TMH1 are important for the signalling process of the TSHR and the LHCGR. The signalling capability is abolished by mutations. The aspartate 410 asparagine substitution is an inactivating pathogenic mutation at the TSHR (de Roux N 1996) that does not influence the hormone binding capability.

Disulfide bonds are involved in the quaternary structure of the TSHR as described in section *1.2.1.1.3.1.* Taken together, the two neighbouring cysteines C283 and C284 of C-b2 are very likely paired either to C398 and C408 of C-b3 or in reverse order to C408 and C398, whereas the C301 (C-b2) is very likely paired with C390 (C-b3).

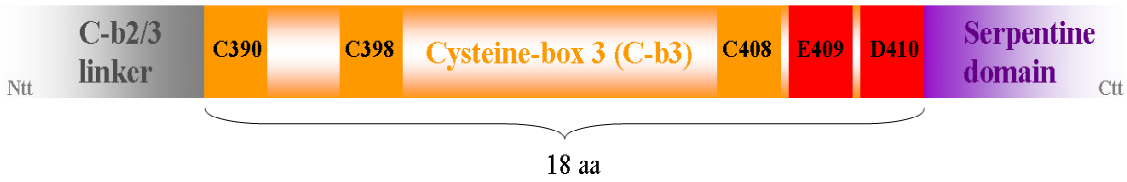

Figure 1.8: *Cysteine-box 3 of hTSHR*
The cysteine-box 3 follows the cysteine-box 2/3 linker region back-to-back as an integral part of the hinge-region close to TMH1. The three cysteines are most likely disulfide bridged to cysteines at cysteine-box 2 and essential for receptor folding by connecting the extracellular with the transmembranal regions. Pathogenic inactivating mutations at E409 and D410 (red) underline the importance of this region for receptor function. E409 and D410 might be important for the formation of the active receptor conformation by binding to a specific interaction partner after hormone-induced signal initiation. Mutations at these positions are unable to stabilize the active receptor structurally and/or functionally.

1.2.1.2 Extracellular loops

1.2.1.2.1 Extracellular loop 1

Mutagenesis studies on the hFSHR and hLHCGR have shown, that the second and third ECLs are essential in GPHR signal transduction (Ryu KS 1996, Fernandez LM 1996 (a), Fernandez LM 1996 (b), Ryu KS 1998, Li S 2001, Sohn J 2003). In contrast to ECLs 2 and 3, little is known about the function of ECL1 in the GPHRs (Ji I 1995) but the characteristics of the TSHR *in vivo* mutations I486F,M (constitutively activating) (Parma J 1995, van Sande 1995, Camacho P 2000) at ECL1 indicate that this ECL has an important role in TSHR signaling.

1.2.1.2.2 Extracellular loop 2

Numerous data for different GPCRs have shown the importance of ECL2 for receptor signalling (Zoffmann S 2002, Seong JY 2003, Sudo S 2003 (b), Herold CL 2004). The ECL2 of the LHCGR has been reported to be involved in hormone binding (Couture L 1996, Ryu K 1998) and to be essential in signal transmission processes (Li S 2001).

For the ECL2 of the TSHR only little experimental information is available, resulting from functionally characterized pathogenic activating mutants I568T,V (Parma J 1995, Claus M 2005). Most pathogenic CAMs, which are reported for the ECLs of the TSHR, involve hydrophobic amino acids (ECL1: I486M,F (Parma J 1995); ECL2: I568T,V (Parma J 1995, Claus M 2005); ECL3: N650Y, V656F (Tonacchera M 1995, Fuhrer D 1997). Disruption of hydrophobic interactions by mutations at these positions might be responsible for structural rearrangements resulting in constitutive receptor activation. However, the detailed mechanisms or interaction partners of these signalling sensitive amino acids have not yet been identified. The main reason is the lack of structural information that could provide insights into intramolecular interactions inside the SD or between the SD and ECD.

1.2.1.2.3 Extracellular loop 3

Three different constitutively activating *in vivo* mutations in ECL3 (Tonacchera M 1995, Fuhrer D 1997, Wonerow P 2000, Claus M 2005) point towards the functional importance of this loop. Mutagenesis at the homologous FSHR and LHCGR has shown that the ECL3 plays an important role in signal transduction (Fernandez LM 1996, Ryu KS 1996, Li S 2001, Sohn J 2002, Sudo S 2003). Alanine-scanning mutagenesis in the LHCGR ECL3 revealed that with the exception of P575 and V579 (equivalent to P652 and V656 in the TSHR) mutations caused only moderate effects on LHCGR binding and signalling (Ryu KS 1996). In a comparable approach, different results were obtained for ECL3 of the FSHR. All mutations resulted in a decrease or loss of Gαs mediated signal transduction and influenced FSHR cell surface expression and ligand binding. Moreover, introduction of alanine at positions L583 or I584 (equivalent to L653 and I654 in the TSHR) caused improved affinity for FSH, while cAMP signalling was abolished (Ryu KS 1998 (b), Sohn J 2002). The homologous positions of K660 in the TSHR were shown to be essential for signalling in the FSHR (K590) and LHCGR (K583) (Fernandez LM 1996, Gilchrist RL 1996).

1.2.1.3 TSH - TSHR interaction

The crystal structure of the FSHR LRRD-FSH complex (Fan QR 2005 (a)) provides insights into the glycoprotein hormone recognition and the binding mode at the extracellular LRRD of a GPHR class member (Fan QR 2005 (a) (b)) (Figure 1.9B). Due to high sequence homology between the TSHR and the FSHR (Vassart G 2004 (a)) and supporting experimental findings (Costagliola S 2003) the general orientation and location of TSH at the LRRD of the TSHR can be assumed to be similar to that established for the FSHR LRRD/FSH complex.

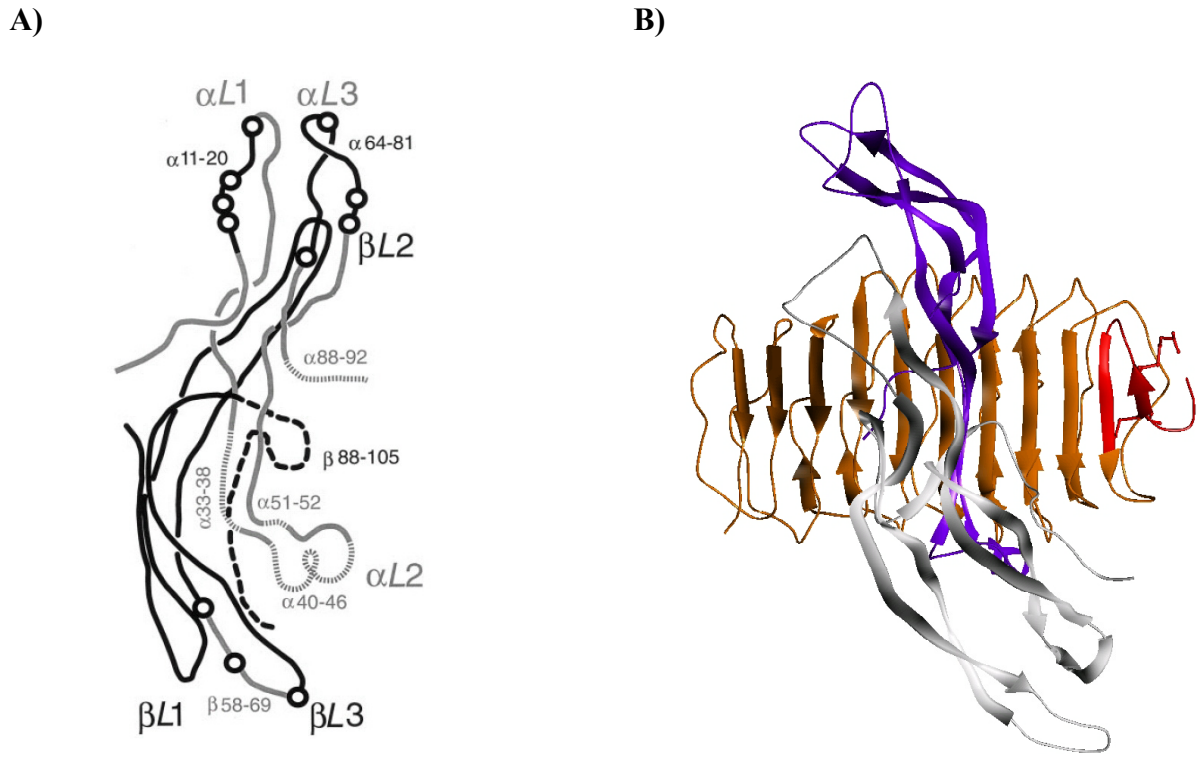

Figure 1.9: ***Binding of the hormone to the Leucine-rich repeat domain***
A) Scheme of glycoprotein hormone-architecture. B) Crystal structure of the FSHR LRR domain complexed with FSH. The α- (purple in B) and β-subunit (white in B) of the hormone are both involved in the binding to the LRR hormone binding domain (orange in B).

The receptor wraps around the middle section of the hormone molecule and interacts with the C-terminal segments of both the FSH α and β-subunits, as well as the αL2 and βL3 loops (Figure 1.9). The hormone is oriented such that loops βL1 and βL3 extend out from the C-terminal tips of the receptor's inner sheet, whereas loops αL1 and αL3 protrude away from the base. The αL2 loop with its helical segment clamps over the top of the LRRD and contacts loops adjacent to the β-sheet. Therefore, both FSH subunits are involved in the intermolecular interaction with the LRRD. The resulting large LRRD/hormone interface has a high charge density. However, the X-ray structure for LRRD of the extracellular FSHR fragment does not

explain all known hormone binding data, for instance, that the introduction of basic residues into the α and β loops of TSH turns TSH into a super agonist (Szkudlinski MW 2004). This indicates that further domains might be involved in the interaction.

Besides the binding mode of GPH to the LRRDs of GPHRs little is known about the first steps of initial receptor activation induced by hormone binding. Only two amino acids of the extracellular domain were identified as directly involved in the receptor activation by *in vitro* studies (Costagliola S 2002). Negative charges in the C-terminal part of the hinge region might play an important role for ligand binding and receptor activation for all GPHRs. This confirms reciprocal findings at the GPHs, where mainly positively charged amino acids are involved in signal transmisson (Zeng S 1999, Hong S 1999, Szkudlinski MW 2004). However, the detailed interrelation between charged residues of the TSHR and TSH are unknown.

One of the most interesting questions is the mechanism/sequence of events for initial receptor activation followed by signal transduction through the TMH region, and the G-protein coupling. The main hypotheses are 1) the disruption of postulated intramolecular silencing interactions between the ECD and SD (not yet identified) (Zhang ML 1995, Vlaeminck-Guillem 2002, Karges B 2005, Neumann S 2005) followed by receptor activation and 2) the receptor activation is initiated by the release of an intramolecular agonist that is activated by mutation or hormone action (Sangkuhl K 2002).

The molecular influence of observed dimerization of GPHRs on the activation mechanism including G-protein activation is still unclear but under intensive investigation (Graves PN 1995, Graves PN 1996, Osuga Y 1997, Latif R 2002, Ji I 2002, Tao YX 2004, Ji I 2004, Urizar E 2005 (b)).

1.2.1.4 Molecular Models of the GPHRs

1.2.1.4.1 LRR domain

All homology models of the LRRD for GPHRs, which were generated prior to 2004 (Moyle WR 1995, Jiang X 1995, Bhownick N 1996, Smits G 2002, 2003), were (almost) exclusively based on the LRRD X-ray structure of the ribonuclease inhibitor (RI) protein (Kobe B 1993) (~15% sequence similarity to the LRRDs of GPHR) showing a distinct horseshoe shape and helices on the convex side for backbone stabilization. The major differences between the predicted concave hormone binding sites and the FSHR X-ray structure published in 2005 (Fan QR 2005 (a)) were the much smaller radii forming the arch, caused by bulky helices on

the convex side that occur in the RI template. It is also the case that the backbone stabilization of the LRRD is not provided by helix-helix interactions between the single leucine-rich repeats, but rather by aromatic interactions. Moreover, the crystal structure of FSHR LRRD provides clear evidence that the cysteine-box 1 is forming an additional LRR as an integral and fold-stabilizing N-terminal part of the LRRD.

1.2.1.4.2 Receptor models including extracellular loops and the hinge region

Complete TSH receptor models where generated by Miguel and coworkers (Miguel RN 2004), and for the homologous LHCGR by Moyle et al. (Moyle WR 2004). For the serpentine domain models the rhodopsin X-ray structure (Palczewski K 2000) is commonly used as a template.

For the LRRD and the hinge region Moyle et al. used the LRRD of the *ScfUbiquitin ligase, SKP1-SKP2* complex (pdb entry code 1FQV) and the *metal-binding domain of the menkes copper-transporting ATPase* (pdb entry code: 2AW0), respectively. Miguel et al. favoured the RI LRRD (pdb entry code 2BNH) for the LRRD modelling and the hinge region was modeled based on the *human tissue inhibitor of metalloproteinase*-2 (*TIMP-2)* template (pdb entry code 2TMP). The molecular models aimed to assemble the different receptor components using experimental findings from mutagenesis, peptide studies, and antibody-mapping. However, these hypotheses have not been proven experimentally, and finally the FSHR/FSH crystal structure (Fan QR 2005 (a)) overruled the theoretical models.

1.3 Activation and Inactivation of the TSHR

1.3.1 Activity states of the TSHR

In this section the current knowledge regarding different activity states will be discussed.

Investigation of relationships between structure and function based on analysis and comparison of mutant phenotypes is a useful strategy to identify epitopes or cores of high functional importance (Schoneberg T 2002). This information is a prerequisite for understanding molecular mechanisms like signal initiation, intramolecular signal triggering and transmission, intermolecular signal transduction to the G-protein, as well as molecular reasons for protein-dysfunctions occurring in mutation-caused diseases (Schöneberg T 2004, Matsushima N 2005).

The basal activity of the TSHR is dependent on the cell type and assay that was used for receptor characterization, but in general a basal level between 5-10% compared to full activity

induced by TSH can be observed (Figure 1.10). Activity levels caused by constitutively activating mutations are very diverse. Mutated receptors with more than 150% of basal TSHR activity are defined as constitutively active.

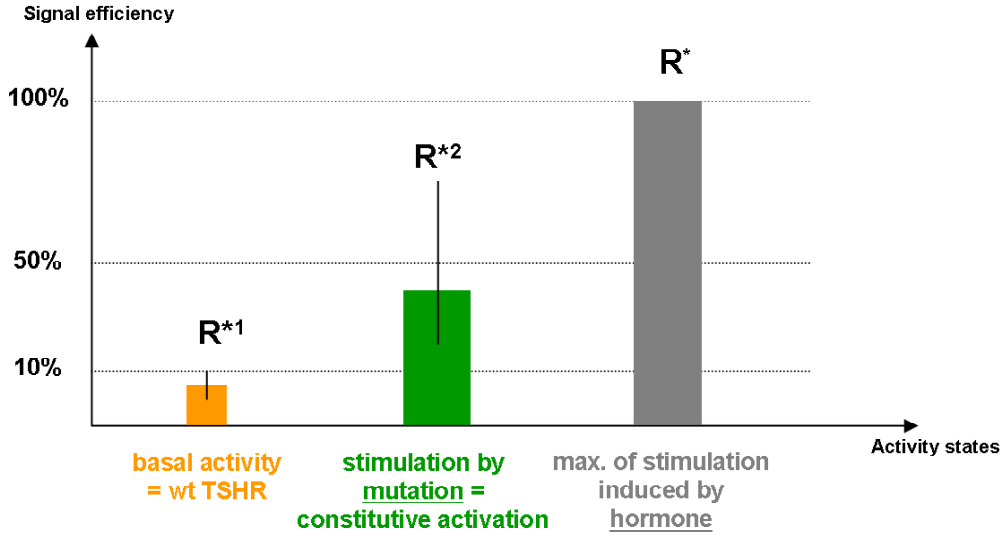

Figure 1.10: ***Comparison of efficiency levels of different activity states***
The fully stimulated TSHR by TSH (R) represents 100% of signalling activity. In comparison, the basal activity level of signalling is between 5-10%. The level of constitutive activation caused by mutations is variable from 20% - 80% compared to full stimulation.*

The levels of signalling activity of the TSHR revealed by *in vitro* mutation analyses are different. In figure 1.11 (Kleinau G 2006 (b)) the functional information resulting from experimental analysis of receptor phenotypes is reflected. Several mutations or TSH action lead to different activity states of the TSHR.

The global difference (delta) between these events is the activation or inactivation of the TSHR. Phenotypes represent distinct levels of corresponding activity states of the receptor that can be distinguished by functional characterization, mainly by the analysis of second messenger accumulation like cAMP or IP.

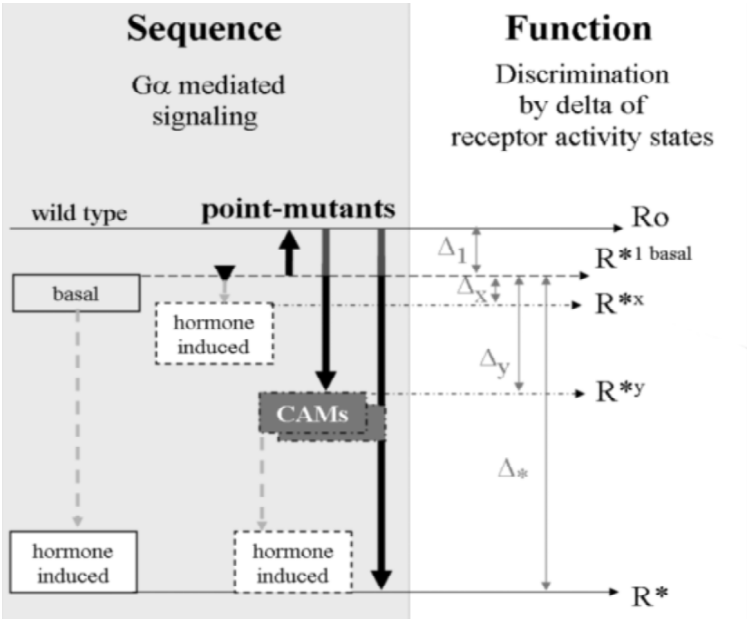

Figure 1.11: ***Relation between receptor phenotypes and activity states of the TSHR*** *(Kleinau G 2006 (b))*
Different receptor phenotypes that are generated by mutations (black bold arrows) like constitutively activating mutants (CAMs) (grey dashed arrows) or the receptor activated by the hormone represent levels of different activation states of the receptor (Ro: abolished basal activity; $R^{*1basal}$: basal activity, $R^{*x...y}$: intermediate active states, R: complete induced activation) (function of the protein).*

1.3.2 Phenotypes of Mutations

Experimentally identified inactivating and activating mutations of GPHRs provide clues about the involvement of particular residues in establishing conformations, which define different receptor activity states (Ro to R*; Figure 1.11) (Cotecchia S 2003, Kleinau G 2004, Kristiansen K 2004, Urizar E 2005 (a), Puett D 2005, Zhang M 2005, Vauquelin G 2005).

1.3.2.1 Constitutively activating mutations

CAMs increase the activity of the receptor for G-protein activation despite interaction with the ligand and they are known for all three GPHR subclasses (Seifert R 2002). Although most known CAMs were identified by characterization of pathogenic mutations (hyperthyroidism), some were also identified by site directed mutagenesis (Kleinau G 2004). The majority of CAMs are located in the SD and they lead to a shift in the equilibrium between the activity states towards the active state. It can be assumed that specific substitutions at these signalling sensitive positions leads to destabilization of the basally active state resulting in receptor activation.

For extracellular located mutations at S281 (hinge region) a special mechanism of constitutive activation is intensively discussed (Sangkuhl K 2002, Vlaeminck-Guillem V 2002) (see also section *1.2.1.1.3.1*). Two opposing hypotheses concerning receptor activation via S281X mutations have been suggsted. One of these postulates that the activation takes place via release of an intramolecular agonist (Sangkuhl K 2002 for LHCGR), and the other that the release of an intramolecular inverse agonist silences the receptor in the basal state

(Vlaeminck-Guillem V 2002 for the TSHR). At present neither of these hypotheses can be favoured above the other.

1.3.2.3 Inactivating mutations

Different mechanisms of complete or partial receptor inactivation caused by mutations (single or multiple substitutions, or deletions) are known: 1) disruption of protein trafficking (Nechamen CA 2000); 2) incomplete or changed protein folding leads to a non-cell surface expressed receptor (Kosugi S 1996 (b)); 3) decreased hormone binding (Smits G 2002); 4) a cascade of constituents involved in the signal transmission process is interrupted via breaking of hydrogen bonds, of electrostatic salt bridge interactions or hydrophobic interactions (Urizar E 2005 (a), Fricke-Otto S 2005); 5) formation of the active receptor state conformation by binding to a new interaction partner after hormone-induced signal initiation is interrupted – the mutant is unable to stabilize the active receptor structurally and/or functionally (Huang JD 1995); 6) disruption of the receptor / G-protein interaction after receptor activation by modification of recognition patterns at the receptor (Claus M 2006). These types of receptor inactivation caused by mutations can be characterized by a variety of phenotypes: i) e.g. the receptor is not observable at the cell surface (possible reason: trafficking interrupted, protein folding disturbed); ii) the ligand binding capability is decreased (possible reason: binding sensitive amino acid has been mutated); iii) or the cAMP/IP accumulation is impaired (possible reasons: ligand binding impaired, receptor membrane trafficking is interrupted, signaling cascade is interrupted).

1.3.2.3 Databases of mutation phenotypes

Known databases of mutations of GPCRs are the *G Protein-Coupled Receptor Data Base* (GPCRdb) (Horn F 1998), the tiny-GRAP database (Kristiansen K 1996), and a database of pathogenic TSHR mutations (University of Leipzig, www.uni-leipzig.de/innere/TSH). These databases collect information concerning several types of mutations or functional studies (point-substitutions, deletions, mutagenesis studies, *in vivo* mutations etc.) in a non-quantitative manner (text based). While these databases are useful informative resources, they do not provide tools for mutant phenotype analyses for the investigation of structure-function relationships, e.g. focused filtering of special mutant-characteristics, or 3D visualization to delineate 'hot-spots' of signalling.

1.3.2.4 Modulation of TSHR activity by low molecular weight ligands

The pathogenic mechanisms of both of the two most common forms of hyperthyroidism are associated with the TSHR. Activating somatic TSHR mutations are one molecular cause of toxic thyroid nodules and toxic multinodular goiter (Wonerow P 2001). The TSHR is also a target for activating antibodies in Graves' disease (Mc Lachlan SM 2005). Antithyroid drug therapy with thionamides is associated with a high rate of adverse reactions and in the case of toxic thyroid nodules it is also directed against the non diseased thyroid tissue (Miehle K 2003). Stimulation of the TSHR by injected recombinant human TSH (rhTSH) is used clinically to promote radioiodine uptake and thyroglobulin secretion during surveillance of thyroid cancer patients and allows symptoms of hypothyroidism to be avoided (Woodmansee WW 2004). Therefore, orally active low molecular weight (LMW) antagonists may have therapeutic potential for the TSHR-mediated hyperthyroidism and LMW agonists might replace injected rhTSH in the diagnosis of thyroid cancer. Up to now, no low molecular weight ligand has been found to be a modulator of TSHR activity.

1.4 Aims of the study

In this section the aims of the studies will be described with respect to initial motivations.

Investigation of the structure-function relationships in the signalling processes of the TSHR is of high importance so as to understand the molecular causes of frequently occurring receptor dysfunctions produced by pathogenic mutations and antibodies. Moreover, knowledge about structural localization and biochemical/physical properties of signalling sensitive amino acids or local areas regarding inactivation and activation of the TSHR is not only necessary for the understanding of the complete activation mechanism but is also helpful for the development of new ideas and sites for pharmacological interventions.

In the last 15 years the knowledge with respect to the intra- and intermolecular functionalities of the TSHR has grown exponentially (focused topics: hormone-receptor interaction, signal transmisson between ECD and SD), but the lack of structural information (3D structure) hampers the understanding of functional findings revealed by biochemical, genetic, and biophysical experiments (e.g. mutagenesis approaches, fluorescence experiments, antibody binding, chimeric studies, hormone binding studies). Mechanisms such as TSHR signalling, activity-regulation, intra- and intermolecular interactions are still not understood in detail. This was the motivation to design the projects described here, so to get deeper insights in the molecular mechanisms of signal transduction of the TSHR with focus on extracellular

components and processes. Our combinatorial strategy uses bioinformatics and molecular modelling to predict relevant intramolecular structure-function relationships (done at the FMP). These computational approaches guide site-directed mutagenesis approaches to gain experimental proofs (mutagenesis done by co-operation partners from the University of Leipzig, III. Medical Department).

The general aim of this study: is the identification of signalling-relevant molecular determinants and their structural and functional interplay in transmitting the signal after hormone-induced activation of the ectodomain into the transmembrane region. Additionaly, the activation of TSHR initiated by low molecular weight ligands, which bind allosterically is investigated (in co-operation with the NIH, NIDDK, Bethesda, MD, US).

The detailed aims of our studies are:

i) Structural and functional description of determinants that are prerequisites for signal transduction in the extracellular domain,
ii) Identification and description of signalling mechanisms,
iii) Insights into the determination of G-protein selectivity in the signalling process that occurs at the ECD,
iv) Description of binding modes for non-peptide ligands,
v.) Generation of a Sequence-Structure-Function Analysis resource (for phenotypes of GPHRs).

1.5 Strategy

The general backbone of the strategy is an iterative sequence-structure-function analysis starting from sequence comparison and homology modeling of 3D-structures of the TSHR (and other GPHRs) combined with data collected from mutation phenotype analysis to reveal structure-function hypotheses (Figure 1.12).

In detail, the studies focus on the functional-structural interplay between the hinge region and the extracellular loops. Molecular modeling was begun with the SD of the TSHR including the ECLs 1, 2 and 3 based on the rhodopsin structure (comparative modeling), the only one available X-ray structure of a GPCR receptor class A.

For the components of the hinge region (see section *1.2.1.1*, LRRD, C-b1, 2, 3; C-b2/3 linker) homologous structural templates were searched for and used for TSHR models to reveal

insights into the local epitope or domain fold and to develop hypotheses about intramolecular interactions.

Initially, the structural models were combined with published functional data from the TSHR and the homologous FSHR and LHCGR. Therefore, the development of a database-resource to summarize and to compare available published experimental findings for the three GPHRs subtypes at once and to utilize it for sequence-structure analysis was also included. To confirm the initial models, derived potential interactions or amino acids of functional importance were suggested and tested by mutagenesis (co-operation partner from the University of Leipzig). The information (also negative results) was used to refine the 3D models and to develop further functional models.

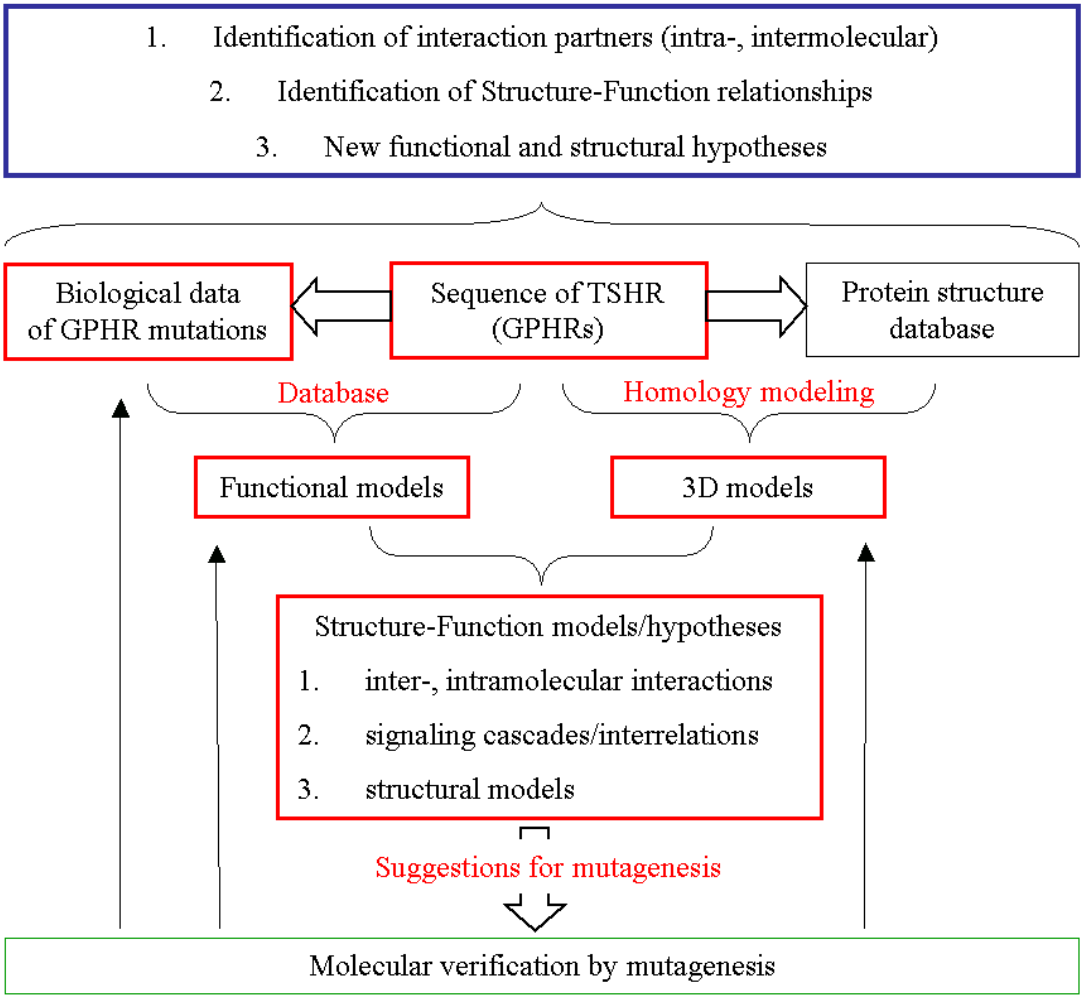

Figure 1.12: ***Working strategy***
The working strategy visualised as a workflow starting from the molecular modeling of 3D-structures of the TSHR (and other GPHRs) combined with data from mutation phenotype analysis to reveal structure-function hypotheses. Red bordered are the methods and responsibilities of the FMP, Berlin, green marked the molecular design and characterization of mutations by the cooperation partner from the III. Medical Department of the University of Leipzig, Endocrinology.

Because the high sequence (structure) homology between the TSHR and the gonadotrophic hormone receptors FSHR and LHCGR, respectively the GPHs, it can be assumed that the results of TSHR-studies could be also valid for the FSHR and the LHCGR, or they can provide molecular insights into differences between the receptor sub-classes.