

2 MATERIAL AND METHODS

2.1 Bioinformatics, Molecular modelling and Docking procedures

2.1.1 Amino acid sequence alignment of the GPHRs

The multiple amino acid sequence alignment of the available 35 GPHR sub-classes was completed by combining automatic multiple sequence alignment techniques using 'Clustal W' (Thompson JD 1994) with a manual refinement (e.g. no gaps allowed within the TM region according to the X-ray structure of rhodopsin (Palczewski K 2000)).

2.1.2 Molecular Modelling of the TSHR and the LHCGR; LMW ligand docking procedures

3D-models were designed based on homology modeling methods for the following receptor components:

1. TSHR and LHCGR serpentine domain
2. Cysteine-box 1 (C-b1)
3. Leucine-rich repeat domain (LRRD)
4. Cysteine-box 2 (C-b2)
- 4a. Aromatic environment of S281 at cysteine-box 2 (C-b2/ECL1)
5. Cysteine-box 3 (C-b3)
6. Extracellular loop 2 (ECL2)
7. Extracellular loop 3 (ECL3)
8. Docking models of the TSHR and the LHCGR in complex with a LMW ligand

In this section, utilized bioinformatics procedures will be summarized and specificities for each study and structural models will be described in detail.

2.1.2.1 The serpentine domain of the TSHR and the LHCGR

The initial 3D structure of the serpentine domain of the TSHR and the LHCGR was established on the basis of the 3D structure of bovine rhodopsin (Palczewski K 2000, Li J 2004, Okada T 2004) (PDB entry codes: 1F88, 1HZX, 1L9H). Several receptor-specific corrections for the GPHRs models were made based on sequence alignments using SeqLab (Wisconsin Package Version 10.2, Accelrys Inc. San Diego, USA).

The designation of the amino acids in the transmembrane domain was based on the Ballesteros-Weinstein nomenclature (Ballesteros JA 1995). In rhodopsin, interactions of the side chains of two consecutive threonines with the helical backbone of the preceding residues caused a structural bulge in TMH2. In the TSHR and the LHCGR, neither consecutive threonines nor prolines exist in indicating a regular α -helix, which extends to residue position 2.71. In TMH5, a minor change of orientation (10 to 15 degrees twist) of the N-terminal half of TMH5 occurs because a proline, that is present in rhodopsin, is absent in the TSHR. Consequently, residue position 5.42 is oriented towards the interior of the receptor.

Loops were added by best fit and homology to fragments of other proteins (from PDB, Berman HM 2000). For the remaining parts of the extracellular loops one and three, conformational fragments of four to seven residues were retrieved from the 3D protein database PDB by means of FASTA (Berman HM 2000). Overlapping fragments with a similar backbone conformation that occur more than once in the database were used for assembling the loops. The sheet-like fold of ECL2 and its general localization were kept as in rhodopsin based on rhodopsin structure-consistent results about different accessibility of two CC chemokine receptor 5 (CCR5) antibodies, each specific for the two different β -strand epitopes of ECL2 of CCR5 (Aarons E.J 2001, Dragic T 1998, Lee B 1999).

All components were modeled with the biopolymer module of the SYBYL program package (TRIPOS Inc. St Louis, MO, USA). Conjugate gradient minimizations were performed until converging at a termination gradient of 0.05 kcal/(mol*Å). For all energy and dynamics calculation the AMBER 7.0 force field was used (Case DA 2002). Molecular dynamic simulations for the TSHR model were performed at 300 K using a periodic boundary box for 2.0 ns, and charges were neutralized by adding chlorine ions. The TSHR and LHCGR models were soaked with water resulting in a tri-phasic water-vacuum-water box (ter Laak AM 1999). Initially the atoms were kept fixed to relax the water during minimization. Later on, the entire system was considered without restraints. The geometrical quality of the models was controlled using the program PROCHECK (Laskowski RA 1993).

2.1.2.2 *Cysteine-box 1 and the leucine-rich repeat hormone binding domain*

To select an optimal structural template for the LRR domains of GPHRs, fourteen different protein structures with LRRD available in the protein structure database (Berman HM 2000) were analysed. The LRRD templates were ranked according the closest number of repeats, residues per repeat and highest sequence similarity compared to TSHR, LHCGR and FSHR

(ranked pdb entries: 1OZN, 1FQV, 2BNH, 1YRG, 1D0B, 1M0Z, 1O6T, 1G9U, 1H6U, 1DCE, 1DS9, 1A9N, 1I00, 1FT8).

Bioinformatic procedure for model refinement: For model building, the Sybyl program package (TRIPOS Inc., St. Louis, MO) and the AMBER 7.0 (Case DA 2002) force field were used. Molecular dynamic simulations for LRR domain models were performed at 300 K for 2.0 ns using a periodic boundary box, and charges were neutralized by adding chlorine ions. The models were soaked in a water box. Initially the atoms were kept fixed to relax the water during minimization. Later on, the entire system was considered without restrains. Conjugate gradient minimizations were performed until converging at a termination gradient of 0.05 kcal/(mol*Å), the AMBER 7.0 force field was used. The geometrical quality of the models was controlled using the program PROCHECK (Laskowski RA 1993).

2.1.2.3 *Cysteine-box 2 and cysteine-box 3*

In order to reveal new homologous structural templates in the ectodomain of the TSHR apart from the LRR motif, extensive systematic sequence similarity searches of fragmented sequence portions of different lengths of the ectodomain were carried out using FASTA within the structure database PDB. The bioinformatics procedure for TSHR model refinement based on identified structural templates was similar to that described in section 2.1.2.2.

2.1.2.4 *S281 at cysteine-box 2 and extracellular loop 1*

The construction of an LRRD ectodomain model for the hTSHR has been described in section 2.1.2.2 and modeling of cysteine-box 2 in section 2.1.2.3. The LRRD, the cysteine-boxes and serpentine domain model were assembled by constrained molecular dynamics (MD) simulations using the biopolymer module of the Sybyl program package (TRIPOS Inc., St Louis, MO, USA).

The assembled ecto-/serpentine domain model of the hTSHR was embedded according its extracellular-, transmembrane and intracellular portions in a tri-phase water-vacuum-water box. Initially, the ectodomain atoms were kept fixed to relax the water during minimization. Later on, the entire system was considered without restraints. Minimizations were also performed until converging at a termination gradient of 0.05 kcal/(mol*Å). MD simulations were performed at 300 K for 1 nsec. For both, the AMBER 7.0 force field was used and the geometric quality of the model was controlled by the PROCHECK software.

2.1.2.5 *Extracellular loop 2 and 3*

The methods of molecular modeling procedures for the TSHR serpentine domain including the ECLs 2 and 3 has been described in section 2.1.2.1. The sheet-like fold of ECL2 and its general localization between the transmembrane helices were kept as in rhodopsin based on rhodopsin structure-consistent results for different accessibility of two CC chemokine receptor 5 (CCR5) antibodies, each specific for the two different β -strand epitopes of ECL2 of CCR5 (Aarons E.J 2001, Dragic T 1998, Lee B 1999).

2.1.2.6 *Docking complexes of the TSHR and the LHCGR with a LMW ligand*

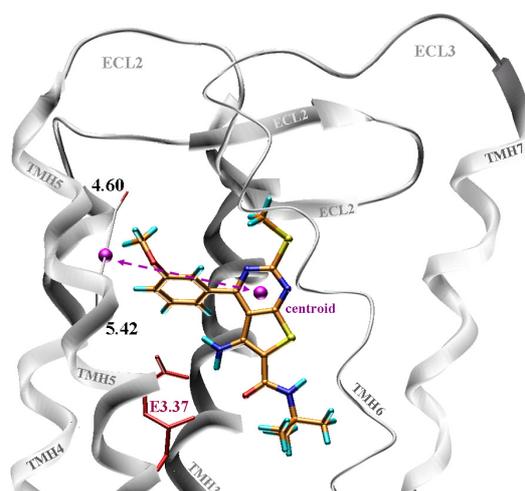
Initially, using the receptor-grid method (Totrov M 1997) implemented in the ICM software (Abagyan R 1994 (a)), the flexible ligands were docked into calculated receptor grids, which reflected the shape, hydrophobicity, hydrogen bonding profile, and electrostatic potential of the receptor. The thirty highest ranked docking-complexes, according to the ICM internal scoring function (Schapira M 1999), were selected in succession for compound docking by Monte Carlo simulation in the internal coordinate space (Abagyan R 1999).

In this second phase, both ligands and binding pocket were treated as flexible. The quality of the complexes was assessed by a scoring function that included grid energy, electrostatics, hydrophobicity, and van der Waals radii (Abagyan R 2001). Numerous high-scoring simulations denoted a hydrogen bond between E3.37 (transmembrane helix 3) and the aromatic amine moiety of active compound org41841, and experimental evidence also suggested such an interaction. Therefore, the investigations were limited to complexes in which the distance between these two groups allowed H-bonding (1-3 Å).

The whole procedure was repeated in quadruplicate for org 41841 at both receptors. Graphic representations have been prepared with the Sybyl package (version 7.1) (Case DA 2002).

Figure 2.1: Distance measuring

The distance of the centroid of the ligand from the medium point of the segment that connects the Ca atoms of the amino acids 4.60 of transmembrane helix 4 and 5.42 of transmembrane helix 5 was calculated. This distance was used to quantify the relative position of the compounds in the different docking poses with respect to transmembrane helices 4 and 5. The Ca atoms of the amino acids 4.60 and 5.42 were chosen based upon their relative orientation within the model and the approximate depth from the extracellular plane as that of the centroid of the bound ligand.



2.1.3 *Sequence-Structure-Function resource*

2.1.3.1 *Data Set*

Utilizing the influence of single side chain substitutions on structure-functional effects, point mutational data are extracted from the literature for the human TSHR, human LHCGR, human FSHR, rat LHCGR, and rat FSHR (mostly used for mutagenesis studies).

The data from the following standard assays for wild type and mutant phenotypes are included: a. cell surface expression level, b. hormone binding capability (maximum), c. basal $G_{\alpha s}$ mediated activity (cAMP accumulation), d. basal $G_{\alpha q}$ mediated activity (IP accumulation), e. $G_{\alpha s}$ mediated activity after hormone treatment (maximum), and f. $G_{\alpha q}$ mediated activity after hormone treatment (maximum). The original data of each publication and assay are scaled to unified percentage values which are calibrated to the corresponding wild type values of 100% and rounded up or down to the closest decimal place. Results that are presented as diagrams in publications are only partially readable for the extraction of absolute values. Therefore, when extracting data from studies presenting results as diagrams the values are strongly rounded or not specified. These mutations are marked in the comment field. In addition to the functional data, experimental conditions such as the cell system used (e.g. COS, HEK) or the type of hormone used for each experiment are also included to enable comparison from the same or different experimental conditions. Additional features are the comparison of corresponding amino acids between all three GPCR sub-classes, the numbering using the three most popular numbering systems, and the citation of the original study. Keywords in a 'comment' field are searchable via the 'Advanced Search', e.g. 'pathogenic mutations' or mutations causing promiscuous hormone binding.

2.1.3.2 *Alignment*

The multiple sequence alignment of human TSHR, human LHCGR, rat LHCGR, human FSHR and rat FSHR is completed by combining automatic multiple sequence alignment techniques using Clustaw ([Thompson JD, 1994](#)) with a manual refinement (e.g. structurally corresponding extracellular cysteines aligned to each other, no gaps allowed within the TM region according to the X-ray structure of rhodopsin). The amino acids are linked to structural features (location, structure, sub-structure).

2.1.3.3 *Numbering*

To achieve unique identification of residues and transparent navigation the available different sequence numbering schemes according to the GPHR alignment are implemented. (i) The sequence for each specific GPHR separately, including the signal peptide (Num1). Published mutagenesis data for FSHR and LHCGR that are numbered excluding the signal peptide sequence are translated into the numbering system including the signal peptide. To facilitate analysis of the comparative relationship between the receptors, two widely spread GPCR residue indexing systems are also offered (ii) the GPCRDB numbering system and (iii) Ballesteros-Weinstein (Ba-We) nomenclature (Ballesteros JA 1995). A highly conserved residue at each helix is used as a common reference for all GPCRs in family A. For example, the highly conserved N at TMH1 is defined as 1.50 and the highly conserved P from the NPxxY motif of TMH7 is defined as 7.50.

2.1.3.4 *3D-models and structural templates for GPHR models*

Bovine rhodopsin, the only currently available X-ray crystal structure (Palczewski K 2000) of 7TMRs, is well established as a structural template for homologous 7TMR structural models (serpentine domain and helix 8). The procedure for generating homologous serpentine domain models for the human TSHR and the human LHCGR and their specific differences were described elsewhere (see chapter 2.1.2.1). Models for rat LHCGR, human FSHR and rat FSHR were generated utilizing the same procedure.

The large extracellular N-terminal ectodomain containing the hormone binding site in the LRRD is common to all GPHRs. An X-ray crystal structure (Fan QR 2005) of the FSHR LRR domain-FSH complex is available as a structural template for the N-terminal ectodomain excluding the hinge region. Waiving the hormone structure, this template was used for LRRD models of the human TSHR, human LHCGR, rat LHCGR and rat FSHR. They were generated using the same procedure as described previously for the modeling of GPHR LRR domains including cysteine-box 1 (Kleinau G 2004).

Since the templates of the models are originally based either on the inactive rhodopsin or on the hormone bound LRRD conformation and do not represent single mutations (or even multiple mutations) or conformations of different activity states (inactive, basal decreased, basal active, partial active, complete active) of the overall receptor, they are shown only as C- α atom trace representations in the wild type receptor model in our system.

2.1.3.5 *Database Technology*

Data is stored in a MySQL database (<http://www.mysql.org>). The web interface was implemented using PHP (<http://www.php.net>). Additionally, functions for internal database administrations have been developed, but access is restricted (for data insertion and updating). The molecular structures/models are displayed using the Jmol software (<http://www.jmol.com>).

2.1.3.6 *Search functions and Output options*

The provided data set and SSFA tools consider various information concerning GPHR phenotypes and allows a user driven search based on: different GPHR subtypes, different sequence numbering systems, specific amino acids, epitopes or structural domains, specified amino acid properties, types of mutation (change of and to specific residue properties), authors of studies, and specifying comments (e.g. pathogenic mutants). One main advantage is given by the possibility of searching and filtering data by a combination of freely adjustable ranges of normalized values of functional characterizations of phenotypes. Each search function is combined with a subset of output formats. The ‘Search Data Set’ is subdivided into three general types of query forms according to initial requests.

The Alignment Search - An alignment of GPHRs is suitable to get a quick overview of GPHR comparison and reveals at which positions/residues mutations are available. Residues where point mutations are known are indicated and data are accessible, while simultaneously providing information about different numbering systems and its general structural and sub structural location. A tabulated (compressed) output includes type of mutation, corresponding positions in the homologous receptors and a link to ‘Details’ that includes available functional details of the characterization.

The Basic Search - retrieves data of mutations and their effects using global or specific queries according to: 1. sequence range, 2. defined residue and 3. selected structural location or element. Information is provided about: a) the different sequence numbering systems, b) general structure and sub-structure location, c) unified values for six assays (if available), d) cell type used, type of hormone, e) authors and f) general comments. Alternative output options are: i) detailed: Complete mutation information including the corresponding positions at homologous GPHRs, ii) data analyzer: tabulated comparison of unified percentage values for all assays (if available) and (optional) easy global discrimination of distant values using a two color classification.

The Advanced Search - provides freely adjustable ranges of normalized values of standard assays for experimental characterization of receptor phenotypes (minimum-maximum range for all assays is 0-9999, no increase (ni) or no signal (ns) is equitable with 0). Fine-tuned studies using a combinatorial analysis of functional data and/or structural locations by a two-step procedure: (i.) Generating a focused data pool from selected functional % range(s) and structural feature(s); (ii.) Combining it in the output data analyzer with an adjustable two class classifier in % ranges of functional data for discrimination of essential features. These sequence-function relationships can be visualized on 3D structural receptor models using the molecule animation program Jmol.

For the 'Advanced Search' different *output* options are provided: 1. *Detailed* - starts with an 'Overview' of available mutants. This overview is linked to analyzed data including separate visualization on 3D models for each retrieved mutation (if the mutation occurs in the SD or LRR); 2. *Data analyzer* - Tabulated (compressed) comparison of unified % values for selected assays and discrimination of distant values by a two-class classification. Highlighting the C-alpha position of the retrieved amino acids as two color-coded balls enables their functional and spatial distinction on the corresponding 3D-models.

2.2 *Characterization of mutant phenotypes*

All methods described in this section were utilized by the cooperation partner from the University of Leipzig, III. Medical Department, Endocrinology.

2.2.1 *Site-directed mutagenesis*

The TSHR mutants were constructed by PCR mutagenesis using the human TSHR plasmid TSHR-pSVL as a template as previously described (Libert F 1989). PCR fragments were digested with BspTI and Eco91I (MBI Fermentas, Vilnius, Lithuania). The obtained fragments were used to replace the corresponding fragments in the wt TSHR-pSVL constructs. Mutated TSHR sequences were verified by dideoxy sequencing with dRhodamine Terminator Cycle Sequencing chemistry (ABI Advanced Biotechnologies, Inc., Columbia, MD).

2.2.2 *Cell culture and transient expression of mutated TSHRs*

COS-7 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % FCS, 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco Life technologies, Paisley, UK) at 37 °C in a humidified 5% CO₂ incubator. Cells were transiently transfected in 12-well plates (1 x 10⁵ cells per well) or 48-well plates (0.25 x 10⁵ cells per well) with 1 µg respective 0.25 µg DNA per well using the GeneJammer® Transfection Reagent (Stratagene, Amsterdam, NL).

2.2.3 *FACS analyses*

The TSH receptor cell surface expression level was quantified on a FACS flow cytometer. Transfected cells were detached from the dishes with 1 mM EDTA and 1 mM EGTA in PBS and transferred into Falcon 2054 tubes. Cells were washed once with PBS containing 0.1 % BSA and 0.1% NaN₃ and then incubated at 4 °C for 1 h with a 1: 200 dilution of a mouse anti human TSHR antibody (2C11, 10 mg/l, Serotec Ltd., Oxford, UK) in the same buffer. Cells were washed twice and incubated at 4 °C for 1 h with a 1 : 200 dilution of fluorescein-conjugated F(ab')₂ rabbit anti mouse IgG (Serotec). Before FACS analysis (FACScan Becton Dickinson and Co., Franklin Lakes, NJ, USA), cells were washed twice and then fixed with 1% paraformaldehyde. Receptor expression was determined by the mean fluorescence intensity (MFI). The wt TSHR was set at 100% and receptor expression of the mutants was calculated according to this. The percentage of signal positive cells corresponds to transfection efficiency, which was approximately 60-70% of viable cells for each mutant.

2.2.4 *cAMP accumulation assay*

For cAMP assays, cells were grown and transfected in 48-well plates. Forty eight hours after transfection, cells were preincubated with serum free DMEM containing 1 mM 3-isobutyl-1-methylxanthine (IBMX) (Sigma Chemical Co., St. Louis, MO, USA) for 20 minutes at 37 °C in a humidified 5% CO₂ incubator. Subsequently, cells were stimulated with 100 mU/ml bTSH (Sigma Chemical Co.) for one hour. Reactions were terminated by aspiration of the medium. The cells were washed once with ice cold PBS and then lysed by addition of 0.1 M HCl. Supernatants were collected and dried. The cAMP content of the cell extracts was determined using the cAMP AlphaScreen™ Assay (PerkinElmer™ Life Sciences, Zaventem, Belgium) according to the manufacturer's instructions.

2.2.5 *Stimulation of inositol phosphate formation*

Forty hours after transfection, cells were incubated with 2 $\mu\text{Ci/ml}$ [myo-3H]inositol (18,6 Ci/mmol, Amersham Pharmacia Biotech, Braunschweig, Germany) for 8 hours. Thereafter, cells were preincubated for 30 min with serum free DMEM containing 10 mM LiCl, without antibiotics. Stimulation by bTSH was performed in the same medium containing 100 mU/ml bTSH (Sigma Chemical Co.) for 1 hour. Intracellular IP levels were determined by anion exchange chromatography as described (Berridge MJ 1983). IP values are expressed as the percentage of radioactivity incorporated from $^3\text{[H]}$ -inositolphosphates (IP1-3) over the sum of radioactivity incorporated in IPs and phosphatidylinositols.

2.2.6 *Linear regression analysis of constitutive activity as a function of TSHR expression (slopes)*

The constitutive activity is expressed as basal cAMP formation as a function of receptor expression determined by ^{125}I -bTSH binding. COS-7 cells were transiently transfected in 24-well plates (0.5 x 10⁵ cells per well) with increasing concentrations of wt or mutant TSHR DNA (50; 100; 150; 200; 250 and 300 ng per well). For radioligand binding assays, cells were incubated in the presence of 160,000-180,000 cpm/ml of ^{125}I -bTSH (BRAHMS Diagnostica, Berlin, Germany) supplemented with 5 mU/ml nonlabeled bTSH (Sigma Chemical Co., St. Louis, MO, USA). For cAMP assays, 48h after transfection, cells were incubated with serum free DMEM containing 1 mM 3-isobutyl-1-methylxanthine (Sigma Chemical Co., St. Louis, MO, USA) for 1h. Cells were washed once with PBS and then lysed using 0.1 N HCl. Supernatants were collected and dried. The cAMP levels were determined using the cAMP AlphaScreen TM Assay (PerkinElmer TM Life Sciences, Zaventem, Belgium) according to the instructions of the manufacturer. Basal cAMP formation as a function of receptor expression was analyzed according to Ballesteros et al. (Ballesteros JA 2001) using the linear regression module of GraphPad Prism 2.01 for Windows.

2.2.7 *Confocal laser scanning microscopy (CLSM)*

HEK 293 cells were grown in Dulbecco's modified Eagle's Medium supplemented with 10% fetal bovine serum, 100 units/ml penicillin, and 100 $\mu\text{g/ml}$ streptomycin (Life Technologies, Inc.) at 37°C in a humidified 5% CO₂ incubator. Cells were seeded on coverslips into 6 well plates (2.5×10^5 HEK 293 cells per well). The cells were incubated 36 hours before transfection with plasmid constructs (2.5 μg DNA/well) containing the coding sequence of the wt or mutated TSHRs. 48h after transfection, coverslips were rinsed twice with ice cold PBS

and fixed with 2 % paraformaldehyde containing 0.1 % Triton X-100 (for permeabilization) for 30 min at 4 °C. After two 5 min wash steps with cold PBS, the cells were incubated with the primary antibody for 1 h at 4 °C. The TSHR was detected using the anti-human TSHR antibody (2C11; Serotec Ltd., Oxford, UK; 1:500 in PBS). The cells were washed two times for 5 min with cold PBS and the primary antibody was detected by incubation with an Alexa-Fluor[®] 488 -conjugated goat anti-mouse secondary antibody (Molecular Probes, Eugene, USA; 1:1000 in PBS) for 1 h at 4 °C. After four final 5 min wash steps, the coverslips were mounted on glass slides. Confocal analysis was performed on a confocal laser scanning system (TCS SP2; Leica, Wetzlar, BRD) attached to a microscope (DM IRBE; Leica, Wetzlar, BRD) with a x100 oil immersion lens (PL Fluotar 1.3; Leica, Wetzlar, BRD). Sections (0.45 µm) were taken, and representative sections corresponding to the middle section of the cells are presented in Figure 3B. After indirect immunofluorescence staining, no specific fluorescence was observed in untransfected HEK 293 cells, or in transfected HEK 293 cells treated only with secondary Alexa-Fluor[®] 488 -conjugated antibody.

2.2.8 *Statistics*

Statistical analysis was carried out by t test using GraphPad Prism 4.03 for Windows (*** p < 0.001 extremely significant; ** p 0.001 to 0.01 very significant; * p 0.01 to 0.05 significant; p > 0.05 not significant).