The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/35801

Please be advised that this information was generated on 2017-07-02 and may be subject to change.
The glycoprotein-hormone receptor information system (GRIS) presents a comprehensive view on all available molecular data for the lutropin/choriogonadotropin receptor, follitropin receptor, and thyrotropin receptor G protein-coupled receptors. It features a mutation database presently containing 696 point mutations, combined with all sequences and the associated homology models. The mutation information was automatically extracted from the literature and manually augmented with respect to constitutivity, surface expression, sensitivity to hormones, and binding affinity. All information in this integrated system is presented in a G protein-coupled receptor specialist-friendly way. A series of interactive tools such as rotamer analysis, mutation prediction, or cavity visualization aids with the design and interpretation of experiments. A universal residue numbering system has been introduced to ease database searches as well as the use of the information in conjunction with literature data from diverse origins. Users can upload new mutations. GRIS is freely accessible at http://gris.ulb.ac.be/.

Molecule Coupled Receptors (GPCRs) are a large superfamily of membrane receptors that are involved in a wide range of signaling activities. They share a tertiary structure characterized by an extracellular amino terminus, followed by seven transmembrane $\alpha$-helices connected by three extracellular and three intracellular loops, with the C terminus in the cytoplasm. This so-called "serpentine domain" forms the basis for the classification of GPCRs into classes displaying sequence identities (1). Class A, or rhodopsin-like GPCRs, form the largest class. Other classes are class B (secretin, glucagon, pituitary adenylate cyclase activating peptide receptors, etc.), class C (calcium sensing and metabotropic glutamate receptors), and a few smaller classes. GPCRs are activated by extracellular ligands. Most class A receptors have a relatively short N-terminal domain and can be activated by small ligands that bind in a cavity between the seven transmembrane helices (Fig. 1) or by peptidic ligands of which a part will be directed in the same cavity (Fig. 1). Class B and C receptors have an extensive extracellular N-terminal domain that is involved in ligand binding (2, 3). For the class B receptors, it has been suggested that the largest part of the peptide agonist forms an $\alpha$-helix that binds to the N-terminal domain of the receptor while the N-terminal residues of the ligand dock between the transmembrane helices (TMs), thereby mimicking a class A receptor's small ligand (2). For the class C receptors, which are constitutive dimers, it is known that the small ligand (e.g. calcium or glutamate) binds to the N-terminal domain of the receptor. The prevailing hypothesis is that activation takes place via conformational changes of the dimeric ectodomain, resulting in an ill-defined effect on the structure of the serpentine dimer (3). Glycoprotein-hormone receptors (GpHRs) belong to class A from an evolutionary point of view (1, 4–6). However, they comprise a large, extracellular, N-terminal domain that, just as in class B and C receptors, is involved in ligand binding (Fig. 1). Upon activation, it is believed that the ectodomain switches from an inverse agonist of the serpentine domain into a full agonist, and initiates signal propagation (7). For the GpHRs,
nothing is firmly known yet about the transduction of this signal (8).

GpHRs consist of three members: the thyrotropin receptor (TSHR), the follicle-stimulating hormone receptor (FSHR), and the luteinizing hormone/choriogonadotropin receptor (LHR/CGR). Their large N-terminal domains consist essentially of leucine-rich repeats (LRRs) (9) that are responsible for selective, high-affinity hormone binding that triggers the signal transduction (10–15).

Many natural gain-of-function mutations of GpHRs have been identified in patients suffering from target tissue autonomy resulting, for the TSHR, in hereditary (16) or congenital (17) toxic thyroid hyperplasia (18) and toxic thyroid adenoma (19) or, for the LHR/CGR, in pseudo-precocious puberty of the male (20). These mutations are mainly observed in the transmembrane portion of the LHR/CGR and the TSHR (19, 21). Analyses of the mutant receptors, using molecular models of GpHRs built using the bovine rhodopsin structure (22, 23) as a template, are yielding information on the mechanism of activation (5, 24, 25). The FSHR seems more resistant to constitutive activation. Only four natural mutants were reported, characterized by persistent male fertility despite pituitary insufficiency (26), and by ovarian hyperstimulation syndrome (25, 27, 28).

Understanding of the structure-function relationships of GpHR subfamily requires that mutation data gathered from one receptor can be interpreted in the light of results obtained with another. Zhang et al. (29) introduced a salt bridge between the helices III and VI in the human LHR/CGR. Mapping a large series of mutations on a homology model, conclusions could be drawn about the importance of interactions between these two helices for LHR/CGR activation. Angelova et al. (30) used sequence alignments to combine mutation data using a schematic model of the generic helix pair VI–VII to study interactions between these two helices from mutation data in multiple receptors. Tao et al. (31) use multiple sequence alignments and a homology model to merge mutation data and to draw conclusions about this same helix pair in the FSHR. Fanelli et al. (32) combined a series of homology models of the LHR/CGR mutated at residue M398 (2.43) and site-directed mutagenesis experiments to study the mechanism of constitutivity induced by mutations at this site.

These examples that shed light on the sequence-structure-function relationships of GpHRs all have in common that extensive integration of heterogeneous data was the key to success. To support our own studies in this field (7, 14, 25, 33), we have designed a GpHR information system. Called glycoprotein-hormone receptor information system (GRIS), it holds all 51 presently known GpHR sequences (classified and aligned) and 696 mutations that were automatically extracted from the literature (34) and manually annotated. Homology models for all TM domains and ectodomains have been built and are fully integrated with the sequence and mutation information. A series of fully interactive user-friendly facilities allow GpHR researchers to intuitively use powerful software that uses the system for the design and interpretation of experiments.

GRIS is freely available at http://gris.ulb.ac.be/. The software and scripts to build GRIS are sufficiently general to be used for other GPCR families, and are available from the authors upon request.

RESULTS AND DISCUSSION

The four basic functions of an information system are browsing, retrieval, query, and inferencing (35).

Browsing

Browsing is important to show the user scientists the context of data, such as similar data for related receptors, different experiments for the same receptor, etc. GRIS uses the hyperlink tables of the GPCR database (GPCRDB) (36) for this purpose, and the information about mutations along with links to SwissProt (37), PubMed (http://www.pubmed.com), etc. adds further browsing facilities. GRIS is also rich in internal hyperlinks to facilitate rapid navigation between mutants, sequences, structure models, and computing facilities. Figure 2 shows a so-called snake-plot for a receptor (38).

The residues that are shown on a white background are hyperlinked to the corresponding muta-
tion information, which is, in turn, hyperlinked to a
three-dimensional (3D) model with the mutated resi-
due indicated (see Fig. 3 for an example). All mu-
tation information is hyperlinked to the correspond-
 ing MuteXt (34) entry in the GPCRDB so that all
relevant sentences from the underlying articles are
rapidly accessible too. GRIS makes extensive use of
the Jmol software (http://www.jmol.org) to allow for
browsing through 3D models of the receptors (see
Figs. 3 and 7 for examples).

Retrieval

Retrieval is important to allow users to evaluate data
and results with in-house software they are familiar
with. For example, GRIS presents all homology mod-
els in the standard PDB format which means that any
in-house molecular visualizer can be used to study the
models beyond the facilities GRIS offers. After design-
ing a mutant, the user can obtain the coordinates of
the mutant structure. Sequences can be downloaded
in both aligned and raw format. All mutant information
can be downloaded as a flat file (human and comput-
er-readable file) for use in other (query) software.

Query

The mutant data can be inspected in many ways. The
snake-like plots (see Fig. 2) present mutants in the format
most commonly used in the GPCR field. The multiple se-
quence alignment that is hyperlinked to the mutation infor-
mation gives the user quick access to mutations of the
same residue in different receptors (see Fig. 4). The most
direct query facility is provided by the query form as shown
in Fig. 5, which allows the user to search for mutations as a
function of species, receptor type, structure element, resi-
due position, and mutation type. The mutation data was
obtained using MuteXt (34) (see Materials and Methods
for a brief description of the MuteXt methodology), and missing
information about the expression level, the degree of con-
stitutive activity obtained with the mutation, and the ligand
binding activity, were manually extracted from the literature
and added to the system. Automated MuteXt updates are
scheduled three times per year. Because it is not possible
to automatically extract all mutant information, we added a
facility that allows users to interactively add mutation infor-
mation to GRIS. Upon reception of a GRIS account, users
can rapidly upload missing GpHR mutations that have been
published in the literature by completing a simple mutation
contribution form (see supplemental data, published on The
Endocrine Society’s Journals Online web site at http://men-
d.endojournals.org). After revision by the curator, the new
mutations are added to the public database. Because Mu-
teXt is an automated literature retrieval method, it has its
limitations. MuteXt needs to access online full-text articles.
The availability of these articles is, in turn, dependent on the
subscriptions of the institution MuteXt is run from. GRIS
users are encouraged to report any missing references.

Three numbering schemes commonly used in the
literature for the TM domains are the GPCRDB stan-
dard (36), the Ballesteros and Weinstein scheme (39),
and the SwissProt numbers (37). All three numbering
systems were implemented into GRIS to allow for
meaningful comparisons of GRIS data and results with
the literature and other external sources of informa-
tion. The location of residues can also be visualized
using publication-ready YASARA (40) pictures (see
Fig. 6).

Inferencing

The most important reason underlying our design of
GRIS is the desire to interpret results and to design
new experiments, especially point mutations that can
shed light on GpHR related questions. For this pur-
pose, a series of WHAT IF (41) mutant design options
has been prepared for output via Jmol. These facilities
include rotamer prediction (42), cavity visualization,
mutant prediction, and the generation of detailed in-
formation about one residue. Figure 7 illustrates some
of these facilities.

The rotamer analysis module of GRIS (Fig. 7A) has
already shown to be an excellent mutation prediction
tool in a recent experiment (43). A constitutively active
M626I (6.37) mutation was found in the TSHR in af-
fected members of a family with nonautoimmune hyperthyroidism. The rotamer analysis using GRIS indicated a severe bump of the mutant residue with I515 (3.46) in the TSHR. We tried rotamer distributions for a few residue types at position I515 in the TSHR-M626I variant. A methionine at position 515 looked most promising and the double mutant I515M/M626I was constructed in the TSHR wild type. Experiments showed that the constitutive activity obtained from the M626I mutation was reduced again to wild-type activity by I515M.

Although GpHRs are activated by their peptide hormone ligands that bind to the extracellular domain, a small organic molecule has previously been designed that acts as a FSHR antagonist (44). Because impairment of FSHR function by mutations can lead to decreased fertility or infertility (45–47), such drugs are becoming promising nonsteroidal contraceptive agents. Rational drug design requires knowledge about cavities in the target receptor, and the identification of residues around such a cavity is obviously important (Fig. 7B).

A previous study by our group stressed the importance of a hydrogen bond network between TM6 and TM7 of the human TSHR to maintain the inactive state of the receptor (24). We reanalyzed all single mutations with the mutation prediction module of GRIS and came to the same conclusions. Mutations D633N
(6.44) and N674A (7.49) still maintain hydrogen bonds between TM6 and TM7 as described in the previous publication (24). GRIS predictions of the substitutions D633A and N674D show that a TM6–TM7 link in this region cannot be maintained, either by total disruption of all hydrogen bonds (D633A) or by electrostatic repulsion (N674D), which is associated with a strong increase in constitutive activity (48).

Another study by our laboratory elucidated a link between constitutive activity and promiscuous activation of the FSHR (25) following the discovery of a mutation D567N (6.30) in a patient with spontaneous ovarian hyperstimulation syndrome. Using molecular modeling, we concluded that this mutation was able to break an ionic interaction between D567 (TM6) and R467 (3.50) (TM3). Inspecting the molecular model of

---

**Fig. 5.** The Mutation Query Form

The mutation query form allows the user to search for mutations according to the receptor, species, structural region, constitutivity, and binding affinity. It is also possible to search for specific substitutions (e.g. Asp→Asn) or to search for mutations at a certain position according to one of the three numbering schemes.

---

**Fig. 6.** High-Quality Cartoon Shot of a Residue and Its Direct Environment

A, Colors of the backbone and residues relate to the secondary structure. The figure shows W494 (4.50) from TM4 hydrogen bonding (yellow dotted line) with N403 (2.45) from TM2 in human FSHR. B, A complete view of the human FSHR serpentine domain colored in rainbow gradient. The same residue W494 (4.50) is selected in gray. Pictures were made with YASARA (http://www.yasara.org).
human FSHR from GRIS, it is easy to see that these residues can indeed interact in the wild-type receptor. Aided by the in silico mutation tool and mutation information of the same residue position in other receptors from GRIS, it is straightforward to design a panel of mutations in the human FSHR that either maintain this interaction or break it, so one can draw firm conclusions from the modeling and experimental data (25).

Each of these examples shows that GRIS is able to predict the effect of mutations and to aid researchers in the GpHR field to rationally set up their experiments. Although the homology models of the GpHR ectodomain based on the human FSHR can be treated as highly reliable models, there are some limitations on the homology models of the transmembrane domain of GpHRS provided by GRIS. Although the 3D coordinates of bovine rhodopsin provide the best template to model the transmembrane domain of GpHRS, it is not the perfect template for several reasons (49). First, sequence analyses show that the opsin family differs very much from the other class A GPCRs. This is particularly true for interhelical loops, where identity is so low that it precludes a reliable alignment. Second, the crystal structure of bovine rhodopsin is an antiparallel and thus unnatural dimer. Third, the available bovine rhodopsin structure is the inactive form of the protein. Predictions on constitutively activating mutations should be made with caution.

When additional class A GPCR structures become available, GRIS will be updated by including novel GpHR homology models along with the current models based on bovine rhodopsin. This should allow for progressive upgrading of present models.

MATERIALS AND METHODS

GRIS has been written in Python (http://www.python.org) and uses CGI scripts for interactive activation of options. Snake-
like plots are produced with the residue-based diagram generator (RbDg) (38). 3D structures are presented with Jmol (http://www.jmol.org). WHAT IF (41) is used for sequence alignments, homology modeling, interactive protein structure analysis, and model predictions. The sequence alignment profile needed by WHAT IF was hand-optimized for GpHR purposes. The conversion of the sequence alignment to colored HTML pages was carried out using the MView program (50).

MuteXt (34) is used to identify and extract single point mutations from literature full-text articles by pattern matching along with names of the concerned GPCR names and organism types. Extracted mutation data are checked by two automated validation filters. The first filter verifies that the mutated amino acids are at the indicated positions in the corresponding GPCR sequence(s). The second filter looks for minimal word distances between mutations, protein names, and organisms to discriminate between possible multiple sequences. Validated mutations are integrated into the GPCRDB (36). GpHR-related mutations are then imported into GRIS. At this stage of reviewing, falsely selected papers (not GpHR related) are removed from the database. Retrieved literature is read by the GRIS database curator and mutation annotations regarding constitutivity, cell surface expression, and ligand binding are added manually.

YASARA (40) is used to produce publication-ready 3D structure figures and to perform energy minimization of the GpHR models. GRIS uses the PostGres (http://www.postgresql.org) relational database system for data storage. GRIS generates hyperlinks from several locations to the GPCRDB for easy access to other forms of data (genomic information, phylogenetic trees, discussion groups, etc.).

All 3D models of GpHRs were constructed using the FSH-FSHR complex structure (9) as template for the ectodomain and the bovine rhodopsin structure (23) as template for the transmembrane domain. For completeness, the hormones were also modeled together with the ectodomains when the hormone sequence was available. Homology models were built using WHAT IF’s standard modeling procedure. Only the structural conserved regions, which could be unambiguously aligned, were modeled. For the ectodomain, these include all nine LRRs and the α- and β-subunits of the hormone. The transmembrane domain includes all seven transmembrane helices and the eighth cytoplasmic helix. Interhelical loops were not modeled. To remove bumps and correct the covalent geometry, all structures were energy-minimized with YASARA (40) by applying the Yamber2 force field, using a 7.86-A force cutoff. After removal of conformational stress by a short steepest descent minimization, the procedure continued by simulated annealing (time step, 2 fsec; atom velocities scaled down by 0.9 every 10th step) until convergence was reached, i.e. no energy improvement was found for 200 steps. The set of energy minimized models and the set of “raw” models (not energy minimized with YASARA) are both available for download from the website. WHAT IF structure integrity check reports are available from the web site as well.

The function of residues in GPCRs is mainly determined by their location (51). Therefore, a mutation of a residue at a certain position in one receptor is likely to have a similar effect as the mutation of a different residue at the equivalent position in another receptor. The structural equivalence can therefore be used to transfer information about mutations in one GpHR to all other GpHRs. To aid this transfer of information, a common residue numbering scheme for all GpHRs has been implemented. For the transmembrane domain, we have implemented the numbering schemes of SwissProt (37), the GPCRDB (36), and Ballesteros and Weinstein (39). Neither the GPCRDB, nor Ballesteros and Weinstein numbered the ectodomains, so we devised a scheme that looks like both these. We used the residue naming of the LRRs by Smits et al. (14) and called the residues in the inner-face β-strands X1-X2-L1-X3-L2-X4-X5, where the side chains of the L1 and L2 residues (mostly leucines, but always hydrophobic) are pointing to the inside of the protein. The X3 residue is mostly hydrophilic and points to the outside. We have chosen to denote the X3 residue of the first LRR as 1050, X3 of the second LRR as 2050, and so on. The numbering within each LRR counts down from the XX50 residue to the previous LRR and up to the next.

The specific residues discussed in this manuscript are numbered according to the SwissProt numbering scheme (37) followed by the Ballesteros and Weinstein numbering (39), e.g. D633N (6.44).

Acknowledgments

We are indebted to Dr. Fabien Campagne (Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, NY) for his time and assistance with the residue-based diagram generator (RbDg), the program to generate the snake-like plots. J.V.D. was supported by a fellowship of Communauté Française de Belgique, Actions de Recherche Concertées.

In memory of Florence Horn, who passed away on July 13th, 2006. You will be forever in our hearts, Flo.

Received January 11, 2006. Accepted March 10, 2006.

Address all correspondence and requests for reprints to: Dr. Gilbert Vassart, Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire, Université Libre de Bruxelles, Campus Erasme, Route de Lennik 808, B-1070 Brussels, Belgium. E-mail: gvassart@ulb.ac.be.

This work was supported by the Belgian Program of Inter-university Poles of Attraction (IUAP/PAI P5/30), initiated by the Belgian State Prime Minister’s Office, Science Policy Programming. This work was also supported by grants from the Fonds de la Recherche Scientifique Médicale, Fonds National de la Recherche Scientifique Science Policy Programming, the LifeSciHealth program of the European Community (Grants LS HB-CT-2003-503337 and LS HB-CT-2004-518167), and Fondation Erasme.

J.V.D., F.H., S.C., G.Vr., and G.Va. have nothing to declare.

REFERENCES

the TSH receptor involves switching of the ectodomain from a tethered inverse agonist to an agonist. Mol Endocrinol 16:736–746
21. Shenker A 2002 Activating mutations of the luteinizing choriongonadotropin receptor in precocious puberty. Receptors Channels 8:3–18
31. Tao YY, Mizrachi D, Segaloff DL 2002 Chimeras of the rat and human FSH receptors (FSHRs) identify residues that permit or suppress transmembrane 6 mutation-induced constitutive activation of the FSHR via rearrangements of hydrophobic interactions between helices 6 and 7. Mol Endocrinol 16:1881–1892
34. Horn F, Lau AL, Cohen FE 2004 Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors. Bioinformatics 20:557–568

Molecular Endocrinology is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.