Glycoprotein Hormone Receptors in the Sea Lamprey *Petromyzon marinus*

Author(s): Mihael Freamat and Stacia A. Sower
Source: Zoological Science, 25(10):1037-1044.
Published By: Zoological Society of Japan
DOI: http://dx.doi.org/10.2108/zsj.25.1037
URL: http://www.bioone.org/doi/full/10.2108/zsj.25.1037

ZOOLOGICAL SCIENCE **25**: 1037–1044 (2008)

Glycoprotein Hormone Receptors in the Sea Lamprey *Petromyzon marinus*

Mihael Freamat and Stacia A Sower*

Department of Biochemistry and Molecular Biology, Rudman Hall, 46 College Road, Durham, NH 03824, USA

Secretion of the pituitary glycoprotein hormones (GpH) follitropin, lutropin, and thyrotropin in vertebrates is the main mechanism by which neuroendocrine signals are propagated at the level of the peripheral glands, gonads and thyroid. Receptors of these hormones (glycoprotein hormone receptors, GpH-R) evolved from a common ancestor through gene duplication and subsequent functional divergence during the split of gnathostomes from their agnathan ancestors. Here we review the properties of two novel receptors closely related to gnathostome GpH-Rs identified in the sea lamprey. Although these are the oldest members of this family of receptors described so far in vertebrates, their overall structural features are remarkably close to their mammalian counterparts. However, they cannot be classified unequivocally as either gonadotropin (FSH-R, LH-R) or as thyrotropin receptors (TSH-R) since they share characteristics with both these groups. This may indicate that lamprey receptors reflect in part properties of the ancestral molecule(s) from which all vertebrate GpH-Rs originated. Molecular phylogenetic relationships among gnathostome GpH-Rs are heavily dependent on the functional domain used in analysis. This suggests large variation in functional constraints acting at the level of different segments of the receptor molecule.

Key words: cyclostomes, sea lamprey, glycoprotein hormone receptor, molecular evolution, functional divergence

INTRODUCTION

The glycoprotein hormone receptors are classified as members of the G-coupled protein receptor (GPCR) superfamily of membrane receptors. They form a distinct subfamily characterized by a large extracellular domain which is half of the total length of the mature protein. Recently, a number of newly identified mammalian and invertebrate receptors have been shown to share the same general molecular organization with the GpH-Rs; therefore, the GpH-R family of proteins has been extended to include these new molecules into a larger class termed leucine rich repeat-containing receptors (LGR) (Hsu et al., 1998; Eriksen et al., 2000; Kudo et al., 2000; Scott et al., 2006; Hoshii et al., 2007; Loy et al., 2008).

The importance of GpH-Rs in mammals is primarily related to control of the development and function of gonadal and thyroid glands via two main endocrine pathways: the hypothalamo-pituitary-gonadal (HPG) axis and hypothalamo-pituitary-thyroid (HPT) axis. These receptors bind the pituitary tropic hormones luteinizing hormone (lutropin, LH), follicle stimulating hormone (follitropin, FSH), and thyrotropin hormone (TSH), which are synthesized and released into the blood stream in response to the action of specific hypothalamic releasing factors (gonadotropin releasing hormone, GnRH and thyrotropin releasing hormone TRH respectively). Once in the bloodstream they travel to the target organs (gonads and thyroid), where they stimulate a cascade of processes that result mainly in the synthesis of steroid and thyroid hormones.

The glycoprotein hormones are dimeric proteins composed of two subunits linked by non-covalent interactions. In vertebrates one subunit (the α chain) is common to all glycoprotein hormones while the second subunit (β) is distinct, so it primarily confers specificity to the interaction with the receptor. In the lamprey, a gonadotropin β subunit cDNA was cloned (Sower et. al, 2006). It is proposed from these studies that lampreys have only one pituitary gonadotropin. Another heterodimeric glycoprotein hormone (after LH, FSH, TSH, and CG, choriogonadotropin) recently discovered in rat (Nakabayashi et al, 2002) was termed thyrostimulin for its ability to stimulate TSH receptors. Its presence in lampreys is also under active investigation in our lab. Glycoprotein hormone receptors act by activation of the cAMP-dependent signal transduction pathways in target cells (Moyle et al., 1975). This results ultimately in steroidogenesis and the secretion of testosterone, estradiol, or progesterone (gonads) or of thyroid hormones (thyroid). An alternate pathway might in some circumstances be the PLC/IP3 signaling pathway (Gudermann et al., 1992).

LH-R, FSH-R, and TSH-R paralogs arose from their common ancestor through gene duplications and evolved into orthologous lineages by speciation. Gene duplication is the general accepted mechanism of increase in the size and complexity of organism genomes during the evolution of vertebrates (Sidow, 1996). Duplications create the material for

* Corresponding author. Phone: +1-603-862-2103;
Fax : +1-603-862-4013;
E-mail: sasower@unh.edu
doi:10.2108/zsj.25.1037

functional diversification of protein families. One copy of the gene keeps the original function, while the other(s) can accumulate mutations, evolve divergently into a protein with new function, or be lost from the genome (Knudsen and Miyamoto, 2001). These evolutionary processes are accompanied by functional divergence of both paralogous and orthologous groups through changes in the amino acid sequences of the receptors in the context of a conserved tertiary (3D) structure (Balaji and Srinivasan, 2007; Bastolla et al., 2006). In the case of GpH-Rs, these processes resulted in functional proteins with a high degree of ligand binding selectivity and tissue expression specificity, playing distinct roles in the endocrine physiology of vertebrates (Grossmann et al., 1997).

The details of this process can only be inferred from the study of existing species, based on the molecular phylogenetic projection of actual GpH-R structures deep in the past of vertebrate lineages and on the understanding of the mechanisms of functional evolution of this class of receptors.

After the advent of the molecular biological methods of cDNA cloning, a large number of glycoprotein hormone receptors have been identified and described. The first characterized were the luteinizing hormone receptors (LH-R) from pig and sheep ovaries (Matsuo et al., 1971; Burgus et al., 1972), followed by their homologs in other mammalian species and jawed vertebrates like birds (You et al., 2000a; You et al., 2000b) and reptiles (Borrelli et al., 2001). The last decade has seen an increase in the number of LGRs, including GpH-Rs described in earlier-evolved vertebrates (fish, reviewed in Kumar and Trant, 2001) and invertebrates such as the sea anemone *Anthopleura elegantissima* (Nothacker and Grimmelikhuijzen, 1993), the fruit fly, *Drosophila melanogaster* (Hauser et al., 1997; Eriksen et al., 2000), and the nematode *Caenorhabditis elegans* (Kudo et al., 2000). Recently, two new glycoprotein hormone receptors (lGpH-R I and II) were described in the agnathan sea lamprey *Petromyzon marinus* (Freamat et al., 2006; Freamat and Sower, 2008). Petromyzontiformes (lampreys) is one of the oldest lineages of vertebrates; its origins date to more than 500 mya in the early history of vertebrates (Gess et al., 2006; Janvier, 2006). Therefore, lGpH-R I and II are the earliest diverged members of this class of vertebrate receptors described to date. Both these receptors are remarkably close in structure to their mammalian homologs. Their tissue expression pattern, however, is much less specific, although increased levels of their transcripts were found in testes (lGpH-R I) and thyroid tissue (lGpH-R II). Both receptors are only marginally activated by the gnathostome glycoprotein hormones.

We think that analysis of their sequence features and their possible functional significance in reference to the gnathostome GpH-Rs may help in understanding the mechanisms involved in functional divergence of this important class of receptors. This review focuses on the identification of the main questions pertaining to the evolutionary relationships between the newly found lamprey receptors and their gnathostome homologs. Understanding the sequence of evolutionary events that resulted in the emergence of various vertebrate GpH-R lineages is important for further investigation and elucidation of the mechanisms of the functional divergence of these genes and their co-evolution and correlated evolution with ligands.

METHODS

Fifty-one GpH-R sequences were downloaded from the NCBI GenBank repository. The list of these sequences (Table 1) includes paralogs from all major vertebrate lineages, the two GpH-R lamprey sequences, and the *Drosophila melanogaster* GpH-R sequence as an outgroup. Alignment and phylogenetic analysis were performed as described in Freamat and Sower (2008). Briefly, glycoprotein hormone receptor protein sequences were aligned with the *muscle* v3.6 (Edgar, 2004), *clustalw* v1.83 (Thompson et al., 1994), *t-coffee* v5.03 (Wallace et al., 2006) and *probcons* v1.1 (Do et al., 2005) multiple sequence alignment programs. The comparative quality scores of the alignments were estimated with the *MUMSA* program (v1.0) (Lassmann and Sonnhammer, 2005). The highest relative accuracy was found for *probcons* output, and therefore the corresponding alignment was used for subsequent analysis.

Regions of the multiple alignment corresponding to the extracellular (ED) and transmembrane (TMD) functional domains of glycoprotein hormone receptors were identified based on the annotations of original GenBank sequence records and refined by comparison with the results of signal peptide, protein motifs, and hydrophobicity analyses of lGpH-R I and II. The highly divergent median region of the SSD was removed from the alignments of the extracellular domain. One hundred bootstrap replicates of the initial gapped alignments were generated with *seqboot* in the PHYLIP package (Felsenstein, 1988), and then maximum likelihood distances were calculated with PHYLIP/*protdist*. The neighbor-joining trees (PHYLIP/*neighbor*) were consolidated into one consensus tree (PHYLIP/*consense*), which was used as the guide tree for maximum-likelihood estimation of branch lengths (PHYLIP/*proml*). Many authors recommend the removal of gap-containing columns and of highly divergent positions from sequence alignments prior to molecular phylogenetic reconstruction. It has been shown that this preliminary treatment of data may improve the accuracy of the phylogenetic signal, especially for DNA alignments of distant sequences (for example, see Castresana, 2000). However, the extracellulular and transmembrane regions we used in calculations were well aligned, so the gaps were not removed as recommended in the documentation of the PHYLIP *protdist* protein distance application (http://evolution.genetics.washington.edu/phylip.html).

The percent identity with lGpH-R I and II protein sequences was calculated for each vertebrate GpH-R for the full coding sequence, as well as separately for each of the following: the extracellular domain (leucine rich domain LRD+signal peptide), the signal specificity domain (SSD), the transmembrane domain (TMD), and the intracellular domain (ID), by using a custom Python routine, $I=Ni/Nt$, where I is the percent identity and Ni is the number of identical residues out of Nt, the total number of non gap-only positions in each pairwise alignment.

LAMPREY GPH-R PROTEIN STRUCTURAL ORGANIZATION

Active glycoprotein hormone receptors are relatively large (70–80 kDa) glycoproteins, partially embedded in the plasma membrane through their transmembrane domain. Structural and functional considerations lead to their organization in four main subunits: the leucine rich repeat domain (LRD), which together with the signal specificity domain (SSD) forms the extracellular segment of the receptor (ED); the transmembrane domain (TMD); and the intracellular tail (ID). The large extracellular, hydrophilic N-terminal end accounts for half the total length of the molecule (about 700

Table 1. Vertebrate glycoprotein hormone receptors and their similarity with lamprey lGpH-R I and II. CDS, coding sequence; LRD, leucine rich repeat domain (without the signal peptide and N-terminal Cys-rich box); SSD, signal specificity domain; TMD, transmembrane domain; ID, intracellular domain. The similarity scores are identity scores. Lengths of domains were calculated based on the probcons multiple sequence alignment and by using the annotations of GenBank entries for reference.

Type, Species	GenBank	CDS			LRD			SSD			TMD			ID		
	Accession number	aa	% I	% II	aa	% I	% II	aa	% I	% II	aa	% I	% II	aa	% I	% II
FSHR *Bos taurus*	L22319	695	42.7	44.6	224	42	43.4	86	19.8	22.1	268	64.9	65.3	69	22.5	31.7
FSHR *Bothrops jararaca*	AY189696	673	42.8	44.9	223	40.3	41.6	66	26.4	16.2	268	65.7	69.4	67	21.7	38.6
FSHR *Cavia porcellus*	AY082514	695	41.6	44.7	224	42.5	44.7	86	18.7	22.8	268	63.1	64.9	69	21.7	32.7
FSHR *Clarias gariepinus*	AJ012647	662	40	39.3	226	35.7	37.7	57	31.7	15.7	268	63.8	64.9	61	15.8	22.8
FSHR *Cynops pyrrhogaster*	AB005587	696	40.9	44.4	224	40.7	41.6	90	18.9	18.6	268	64.9	68.3	67	20.8	33.7
FSHR *Danio rerio*	AY278107	668	40.3	40.1	225	35.7	37.7	59	29.2	15.4	268	65.7	66.8	61	16.7	25.7
FSHR *Equus asinus*	U73659	687	41.5	43.8	224	41.2	43.8	78	20.2	22.1	268	62.3	63.8	69	22.5	31.7
FSHR *Felis catus*	AY521181	695	42.8	44.3	224	42	43.4	86	18.7	22.1	268	65.7	66	69	22.5	29.7
FSHR *Gallus gallus*	D87871	693	41.7	44.6	224	41.2	41.6	86	19.8	19.1	268	64.9	68.3	67	20.8	36.6
FSHR *Homo sapiens*	AY429104	695	41.2	45	224	39.8	44.2	86	18.7	23.5	268	63.8	65.3	69	21.7	32.7
FSHR *Ictalurus punctatus*	AF285182	662	40	39.3	226	35.7	37.7	57	31.7	15.7	268	63.8	64.6	61	15	22.8
FSHR *Macaca fascicularis*	X74454	695	41.4	45.1	224	39.8	43.8	86	18.7	22.8	268	64.2	65.7	69	21.7	34.7
FSHR *Macropus eugenii*	AY082002	694	41.3	44.5	224	41.6	41.6	87	18.5	20.4	268	64.2	66.8	67	20.8	37.6
FSHR *Mesocricetus auratus*	AY509907	694	40.9	43.3	224	42	43.4	85	18.9	19.9	268	62.3	64.6	69	18.3	29.7
FSHR *Mus musculus*	AF095642	692	41.4	42.3	224	41.6	42	85	20	19.1	268	64.2	63.8	67	16.9	27.7
FSHR *Oncorhynchus mykiss*	AF439405	659	39.1	38.2	226	36.6	40.4	57	27	14.2	269	63.6	61.7	53	12.3	17.8
FSHR *Ovis aries*	NM_001009289	695	42.7	44.3	224	42.9	43.4	86	20.9	22.1	268	64.6	65.3	69	20.8	30.7
FSHR *Podarcis sicula*	AJ292553	673	38.8	38	224	39.4	32.7	67	25	16.8	267	59.3	60.8	66	15.8	28.7
FSHR *Rattus norvegicus*	NM_199237	692	41.9	43.2	224	41.6	41.6	85	20	20.6	268	64.2	64.9	67	20.3	30.7
FSHR *Salmo salar*	DQ837298	660	39.3	38.5	226	37	40.8	57	27	14.2	268	64.2	62.7	53	13.2	17.8
FSHR *Sus scrofa*	AF025377	695	42.3	43.7	224	41.6	43.8	86	18.7	20.6	268	64.9	65.3	69	22.5	28.7
GTHRII *Oncorhynchus rhodurus*	AB030005	724	38.7	36.5	228	38.9	36.1	97	17.3	13.1	268	61.2	63.8	75	18.8	14.5
LHR *Bos taurus*	U20504	701	40.4	41.2	225	41.6	41.4	79	23.8	19.7	268	61.9	67.2	76	16.7	20.7
LHR *Callithrix jacchus*	U80673	676	41.3	41.5	225	40.7	42.3	52	27.3	14.8	268	62.3	68.3	76	18.3	23.4
LHR *Clarias gariepinus*	AF324540	710	40.2	37.1	225	42	37.9	86	19.8	15.3	268	63.1	63.4	82	16.5	15.3
LHR *Danio rerio*	AY714133	708	39.7	37.4	225	39.8	36.1	93	17.2	13.8	268	64.6	65.3	75	16.8	16.2
LHR *Gallus gallus*	NM_204936	728	39.8	40	225	40.3	42.7	109	15.8	16	268	66.4	68.3	80	16.9	19.1
LHR *Homo sapiens*	S57793	699	40.2	42.7	225	39.4	44.1	79	22.6	19.7	268	62.3	67.9	76	18.3	21.6
LHR *Ictalurus punctatus*	AF285181	696	40.1	36.9	225	39.4	37.9	78	24.4	16.4	268	60.8	61.9	70	17	16.2
LHR *Mus musculus*	NM_013582	700	40.9	41.1	225	41.2	41.9	79	23.8	19.7	268	62.7	67.5	73	18.3	21.3
LHR *Oncorhynchus mykiss*	AF439404	727	38.7	36.5	228	39.3	36.1	100	17	13.1	268	61.2	63.8	75	18.8	14.5
LHR *Rattus norvegicus*	NM_012978	700	40.9	41.3	225	40.7	42.3	79	25	20.4	268	62.7	67.9	73	19.2	19.4
LHR *Salmo salar*	DQ837299	728	38.9	36.5	228	39.3	36.1	100	17	13.1	268	61.2	63.8	75	20.5	14.5
LHR *Sus scrofa*	NM_214449	696	41.2	42.1	225	41.6	43.2	79	25	19.7	268	62.7	67.9	73	16.9	20
TSHRB *Oncorhynchus rhodurus*	AB030955	793	41.5	40.6	226	44.7	41	141	15.6	21.1	268	65.7	69	96	26.5	19.1
TSHR *Bos taurus*	NM_174206	763	42	45.8	226	46	43.6	129	14.7	27.5	268	66.4	72.8	86	21.6	23.7
TSHR *Canis familiaris*	NM_001003285	764	42.9	45.9	226	46.9	44.9	130	15.4	27.3	268	66.8	72.4	86	24	23.7
TSHR *Clarias gariepinus*	AY129556	777	38.8	39.5	226	43.8	38.3	130	16.2	22.5	268	62.3	65.3	100	17.7	18.9
TSHR *Felis catus*	AF218264	763	42.5	46.1	226	46.5	45.8	129	15.5	25.3	268	66.4	72.4	86	23.2	24.6
TSHR *Gallus gallus*	AB234613	761	42.2	46.4	226	46.5	43.6	131	18.3	27.8	268	64.6	75	84	22.8	22.7
TSHR *Homo sapiens*	AY429111	764	42.6	45.9	226	46	44.9	130	16.2	25.3	268	66.8	73.5	86	22.4	22.9
TSHR *Ictalurus punctatus*	AY533543	778	38.5	39.5	226	41.6	38.8	131	17.6	23.2	268	61.9	66	100	18.4	17.4
TSHR *Morone saxatilis*	AF239761	779	42.6	42.2	226	44.2	44.1	138	15.2	20.9	268	66.4	69.8	92	30.1	20.5
TSHR *Mus musculus*	NM_011648	764	42.4	46.4	226	46.5	44.5	130	16.2	28	268	66.8	73.1	86	20.8	22.9
TSHR *Oreochromis niloticus*	AB047390	778	42.3	42.3	226	43.8	42.3	138	15.9	21.5	268	66.8	70.5	92	27.4	21.2
TSHR *Ovis aries*	NM_001009410	764	41.8	45.5	226	46	44.5	130	15.4	25.3	268	65.7	72	86	21.6	23.7
TSHR *Rattus norvegicus*	NM_012888	764	41.9	44.8	226	46.9	43.6	130	16.2	26.7	268	65.3	70.9	86	20	21.2
TSHR *Sus scrofa*	AF338249	764	42.3	46.2	226	45.1	45.4	130	15.4	26.7	268	66.8	73.1	86	23.2	24.6
GpHRI *Petromyzon marinus*	AY750688	719	100	39.9	226	100	42.3	52	100	12.7	268	100	67.9	110	100	19.9
GpHRII *Petromyzon marinus*	AY750689	781	39.9	100	226	42.3	100	132	12.7	100	268	67.9	100	100	19.9	100
GpHR *Drosophila melanogaster*	U47005	831	25.5	26.5	223	24.2	26.3	129	9.2	11.4	272	52.2	50	75	12.5	15.6

I=percent identity with lGpH-R I
II=percent identity with lGpH-R II

amino acid residues on average). Most of it is represented by the Leu-rich repeat domain (LRD), which is flanked at its N-terminal end by a characteristic Cys-rich box followed by nine Leu-rich repeats (Dufau, 1998; Simoni et al., 1997; Ascoli, 2005; Szkudlinski et al., 2002). This motif is found in many proteins usually involved in protein-protein interactions or cell adhesion (Buchanan and Gay, 1996). Its consensus sequence is *LXXLXLXXNXL*, but some deviations from this general pattern exist between the repeats of the same molecule as well as between different paralogs or orthologs. The tridimensional structure of this segment was predicted based on its similarity with other Leu-rich repeat-containing proteins (ribonuclease inhibitor) and was later confirmed by X-ray crystalography in the case of FSH-R (Fox et al., 2001; Fan and Hendrickson, 2005).

Conservation of Leu residues in the Leu-rich domain of lamprey receptors is significantly higher than the overall conservation of the corresponding domain (see Tables 1 and 2). Lamprey GpH-R I has five N-glycosylation sites: three located in the LRD, one in the SSD, and one in the intracellular tail. Lamprey GpH-R II has seven putative N-glycosylation sites, of which four are located in the LRD, two in the SSD, and one in the ID. Some of the lGpH-Rs are very well conserved relative to the gnathostome sequences, while others are unique (Table 3). There are no N-glycosylation motifs in the transmembrane domain of lGpH-Rs, in contrast with the mammalian LH-Rs or fish FSH-Rs. The presence of one glycosylation motif in the intracellular domain is characteristic of mammalian FSH-R, TSHR and some teleost LH-Rs. Both lamprey receptors contain a putative PROSITE (http://ca.expasy.org/prosite/) protein kinase C phorphorylation motif *[TS]-x-[RK]* in the ninth leucine repeat, which is also conserved in most vertebrate GpH-Rs.

The LRD and transmembrane domains are linked by a highly divergent intervening segment. This region is usually characterized as the 'hinge' region and is assigned a simple, merely connective, function; however, there is strong experimental support for the concept of the important role played by this domain in modulation of the selectivity of the receptors towards their ligands, and within this context the linker is termed the signal specificity domain (SSD) (Moyle et al., 2005; Mizutori et al., 2008). The SSD has the lowest overall degree of conservation between GpH-R classes, and even within the same paralogous group the similarity is low. However, the two Cys rich boxes at the N-terminal and C-terminal ends of the SSD are among the best preserved sequence features in all gnathostome and lamprey receptors. The *Y[DE]X* motif located at the N-terminal end of the downstream Cys-rich box was shown to be essential for activation of the mammalian glycoprotein hormone receptors by their cognate ligands (Costagliola et al., 2002) upon sulfation of the tyrosine residue. Both lamprey receptors lack this motif, and in this respect they resemble fish gonadotropin receptors that lost or did not acquire this feature. The length of the lGpH-R I SSD (52 aa) is the smallest of all GpH-Rs, approaching the fish FSH-R value; in contrast, the length of the lGpH-R II/SSD was found to be the closer to the length of the same segment of thyrotropin receptors (ca. 130 aa) (Table 1).

The transmembrane (serpentine) domain (TMD), is the most conserved region of this class of receptors. Overall the number of identical residues in this section region reaches 80% within the GpH-R subfamily. This score is much lower when calculated in respect to members of the the rhodopsin-like group of GPCRs (around 20%), although many residues and motifs characteristic to GPCRs are also present in the GpH-R subfamily. The identity score of both lamprey GpH-Rs varies between 60% and 75% for this region. The TMD contains seven membrane-spanning fragments 20–30 residues in length, on average, interrupted by shorter (7–16 residues) alternating intracellular and extracellular loops. The first and second extracellular loops contain cysteine residues also present in lamprey sequences, and the disulfide link between them is considered important for the stabilization of the transmembrane conformation (Dufau, 1998). The first intracellular loop of lamprey GpH-R I harbors a protein kinase C phosphorylation motif present only in LH-Rs. A similar motif in the second intracellular loop is characteristic only to lGpH-R I, while the third motif at the C-terminal end of transmembrane segment 5 is conserved in thyrotropin and follitropin receptors. In contrast, in lamprey GpH-R II only one PKC phosphorylation motif can be detected, at the C-terminal end of the fifth transmembrane fragment.

Table 2. Conserved Leu residues in the Leu-Rich repeat domain of lamprey GpH-Rs and rat LH-R relative to the consensus sequence of LH-R, FSH-R, and TSH-R. Values were calculated as percent of conserved residues from the total number of Leu residues in each sequence LRD (in parantheses next to each header entry).

	LRD consensus		
LRD	LH-R (34)	FSH-R (29)	TSH-R (34)
Lamprey GpH-R I (32)	68	68	71
Lamprey GpH-R II (34)	67	61	78
Rat LH-R (33)	72	84	63

Table 3. Conservation of N-glycosylation motifs in lamprey GpH-Rs.

motif #	Domain	lGpH-R I	lGpH-R II	LH-R	FSH-R	TSH-R
1	LRD	+	+	all	all	all
2	LRD	–	+	fish	fish	mammals
3	LRD	+	+	all	all	all
4	LRD	+	+	fish	–	–
5	SSD	+	+	all	all	all
6	SSD	–	+	mammals	–	–
7	ID	+	+	–	mammals	all

The transmembrane domain mediates the transfer of signal to the intracellular medium by mechanisms that have not been entirely elucidated. The interaction of the ED/ligand complex with the extracellular loops triggers conformational changes transduced through the membrane to the intracellular medium. The details of this interaction are not well established, and currently there are different models supported by experimental evidence. On the cytoplasmic side, the intracellular loops may be involved in regulatory interactions associated with desensitization of the receptor and/or internalization of the receptor/ligand complex.

The intracellular tail is, together with the SSD, the most divergent part of the GpH-Rs. The N-terminal end of this section, however, is better conserved and contains a very well-preserved Cys residue ca. 20 positions downstream from the last transmembrane segment in both lamprey GpH-Rs. This residue is palmitoylated in vertebrate receptors, and the aliphatic chain of the palmitoyl residue is embedded in the plasma membrane. Removal/replacement of this Cys residue was shown to affect the capability of the receptor to be expressed on the plasma membrane (Lei et al., 2005). Residues between the C-terminal end of the TMD and the palmitoylated Cys form a so-called eighth intracellular loop. Lamprey receptors contain a PKC phosphorylation site in this region that is also present in all FSH-Rs in vertebrates, with the exception of mammals.

MOLECULAR PHYLOGENY OF GLYCOPROTEIN HORMONE RECEPTORS

Analysis of the molecular phylogenetic relationships between the lamprey receptors and the members of the vertebrate GpH-R subfamily (Freamat and Sower, 2008) showed that the former are members of the vertebrate glycoprotein hormone receptor subfamily. The overall topology of the phylogenetic tree has the usual characteristics seen for this class of proteins. There are three groups of orthologous sequences corresponding to LH receptors, FSH receptors, and TSH receptors, respectively. The FSH-Rs and TSH-Rs appear to be monophyletic, and the LH-Rs are in the sister clade. The phylogenetic relationships among these three clades are, however, not clear; the topology seems to be very sensitive to the selection of the taxa used in analysis, and many times it is resolved differently, sometimes even in different papers by the same authors (for example, see Vischer and Bogerd, 2003b; Vischer and Bogerd, 2003a).

The lamprey receptors appear to be clustered as the sister group of the thyrotropin receptor group of orthologs. This is consistent with a general concept that the evolution of vertebrate GpH-Rs was initiated by duplication of an ancestral receptor that was a TSH-R-like homolog. One interpretation of this topology suggests that the duplication that gave rise to lamprey receptors took place exclusively in the lamprey lineage, independently from the series of genome/gene duplications responsible for generating the gnathostome GpH-R paralogous lineages TSH-R, FSH-R, and LH-R. The two copies of the original lamprey GpH-Rs may have undergone subfunctionalization by partition of their tissue expression pattern, while retaining high sequence similarity due to their interaction with a unique ligand.

Examination of the identity scores in Table 1 indicates that lamprey receptors share fewer residues with each other than with the gnathostome GpH-Rs. It is also known that interaction between the receptor and the cognate ligand may induce reciprocal evolutionary constraints usually described in terms of co-evolution of the two interacting partners (Moyle et al., 1994). Experimental evidence accumulated in our lab to date suggests that there is likely only one putative GpH ligand for both lamprey receptors. This would translate to similar evolutionary constraints acting at the level of the extracellular domains responsible for direct interaction with the ligand, and therefore resulting in similar phylogenetic signals. This is why we derived the molecular phylogenies independently for the extracellular domain and the transmembrane domain of vertebrate GpH-Rs. The resulting trees have different topologies (Fig. 1). The tree generated for the extracellular domain reflects more closely the full-sequence tree topology with respect to the position of the lamprey receptors. In contrast, the tree generated for

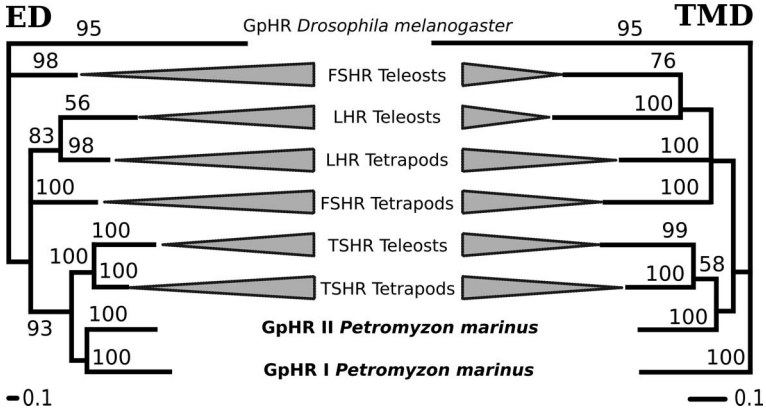

Fig. 1. Comparison of the molecular phylogenies of the glycoprotein hormone receptor extracellular (ED) and transmembrane (TMD) domains. Maximum-likelihood branch lengths were calculated from guide trees representing the consensuses of the neighbor-joining phylogenies calculated from 100 bootstrap replicates for the extracellular domain and transmembrane domain segments of the GpH-R sequence dataset in Table 1.

the transmembrane domain shows the lGpH-R II in the same position as an ortholog of gnathostome TSH-Rs, but places lGpH-R I outside the entire vertebate GpH-R group.

These differences can be understood by considering that different domains evolved with non-equal rates in different phylogenetic lineages as a result of the different evolutionary constraints acting upon the domains. In the case of the vertebrate GpH-Rs, interaction with one ligand (in lamprey) versus three different ligands (in gnathostomes) may require a reevaluation of the methodology of interpretation of molecular phylogenetic data. This is a consequence of the different roles played by GpH-R domains in ligand binding and activation of signal transduction. To actually understand the nature and intensity of selective pressures at different sites, one needs to know the functional roles of the corresponding sites. This in turn requires clarification of the mechanism of ligand binding and receptor activation. Also, it must not be ignored that these mechanisms may have been different for different lineages at different times during evolution.

CONCLUSIONS

The most significant contribution of the study of 'living fossil' organisms like lampreys consists of their capability to substitute to some extent for critical ancestral internal nodes in the tree of life. In molecular studies of GpH-R evolution, this contribution provides more reliable ways to derive the ancestral sequence from which all vertebrate paralog and ortholog lineages have diverged. This is important because we can only study directly the 'evolutionary successes' i.e., the actual proteins which resulted through changes that took place over periods of hundreds of millions of years. The ancient origin of the lamprey lineage and its relative conservation of ancestral features suggest that its study may offer the premises for a more accurate estimation of how the genome, proteome, and signaling networks of the common ancestor of vertebrates may have looked. In the case of the GpH receptors, it is necessary first to establish the precise position of the lamprey receptors in the evolution of this class of proteins.

Fig. 2 shows a diagramatic representation of our hypothesis for the early evolution of the vertebrate glycoprotein hormone receptors and the place of lamprey GpH-Rs in this process. We hypothesize that the first duplication (D1) took place before the divergence of gnathostomes; lamprey lGpH-R I and II paralogs are orthologs of different ganthostome GpH-R groups. The alternate possibility, i.e., that the duplication event which gave rise to lamprey receptors took place within the lamprey lineage, is supported by the topology of the phylogenetic tree derived for the extracellular domain. In our opinion, the evolution of this domain took place under specific constraints induced by interaction with a common ligand and is not reflective of the actual evolutionary origin of these proteins. Analysis of the phylogenetic tree is exposed to risks of misinterpretation, particularly in respect to identification of the duplication nodes as speciation nodes if some paralogs were lost after the duplication events (Seoighe et al., 2003): we did not consider in our interpretation all the numerous alternatives involving duplication/gene loss events compatible with existing information. It is easy to observe that a simple molecular phylogenetic analysis as described above refuses to provide a simple explanation for the relationship of lamprey receptors with their gnathostome homologs. We consider that this is partly due to the modular multi-domain architecture of GpH-R proteins, and in particular to the independent evolutionary dynamics of the extracellular and trans-membrane domains at the divergence point between modern agnathans and gnathostomes. Selective pressures such as changes in the environment may also be different in these domains, ensuring successful reproduction, a central focus for selective agents.

It is likely that analysis of the non-linear relationship between evolutionary rates in these protein domains will provide valuable information for understanding the place of GpH-R I and II in the evolution of vertebrate receptors and of mechanisms of functional divergence of gonadotropin and thyrotropin receptors.

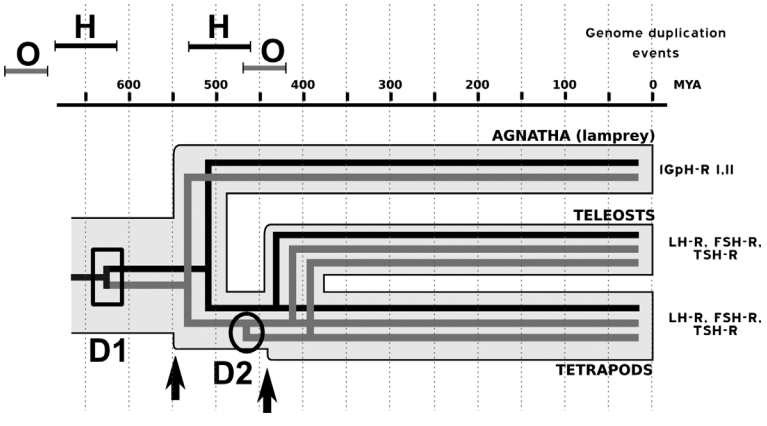

Fig. 2. Early events in the the evolution of glycoprotein hormone receptors in vertebrates. We hypothesize that the gene duplication which gave rise to lGpH-R I and II paralogs (**D1**) took place before the lamprey-gnathostome split. Glycoprotein hormone receptor lineages are represented with solid lines and overlap with the actual phylogeny of extant agnathan, fish, and tetrapod groups shown as a shaded outline. The approximate intervals of gene duplication events are from Holland et al. (1994) (solid black line, **H**) and Ohno (1970) (solid grey line, **O**). The vertical arrows point to the estimated divergence time of gnathostomes and tetrapods, respectively.

ACKNOWLEDGMENTS

This research was supported by NSF grant IBN-0421923. This is scientific contribution number 2363 from the New Hampshire Agriculture Experimentation Station.

REFERENCES

Ascoli M (2005) Learning new tricks from an old dog: the processing of the intracellular precursor of the luteinizing hormone receptor (LHR) into the mature cell-surface LHR is a regulated process. Endocrinology 146: 3221–3223

Balaji S, Srinivasan N (2007) Comparison of sequence-based and structure-based phylogenetic trees of homologous proteins: inferences on protein evolution. J Biosci 32: 83–96

Bastolla U, Porto M, Roman HE, Vendruscolo M (2006) A protein evolution model with independent sites that reproduces site-specific amino acid distributions from the protein data bank. BMC Evol Biol 6: 43

Borrelli L, Stasio RD, Parisi E, Filosa S (2001) Molecular cloning, sequence and expression of follicle-stimulating hormone receptor in the lizard *Podarcis sicula*. Gene 275: 149–156

Buchanan SG, Gay NJ (1996) Structural and functional diversity in the leucine-rich repeat family of proteins. Prog Biophys Mol Biol 65: 1–44

Burgus R, Butcher M, Amoss M, Ling N, Monahan M, Rivier J, Fellows R, Blackwell R, Vale W, Guillemin R (1972) Primary structure of the ovine hypothalamic luteinizing hormone-releasing factor (LRF). Proc Natl Acad Sci USA 69: 278–282

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552

Costagliola S, Panneels V, Bonomi M, Koch J, Many MC, Smits G, Vassart G (2002) Tyrosine sulfation is required for agonist recognition by glycoprotein hormone receptors. EMBO J 21: 504–513

Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S (2005) Probcons: probabilistic consistency-based multiple sequence alignment. Genome Res 15: 330–340

Dufau ML (1998) The luteinizing hormone receptor. Annu Rev Physiol 60: 461–496

Edgar RC (2004) Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797

Eriksen KK, Hauser F, Schiott M, Pedersen KM, Sondergaard L, Grimmelikhuijzen CJ (2000) Molecular cloning, genomic organization, developmental regulation, and a knock-out mutant of a novel leu-rich repeats-containing g protein-coupled receptor (dLGR-2) from *Drosophila melanogaster*. Genome Res 10: 924–938

Fan QR, Hendrickson WA (2005) Structure of human follicle-stimulating hormone in complex with its receptor. Nature 433: 269–277

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22: 521–565

Fox KM, Dias JA, Roey PV (2001) Three-dimensional structure of human follicle-stimulating hormone. Mol Endocrinol 15: 378–389

Freamat M, Sower SA (2008) A sea lamprey glycoprotein hormone receptor similar with gnathostome thyrotropin hormone receptor. J Mol Endocrinol 41: 219–228

Freamat M, Kawauchi H, Nozaki M, Sower SA (2006) Identification and cloning of a glycoprotein hormone receptor from sea lamprey, *Petromyzon marinus*. J Mol Endocrinol 37: 135–146

Gess RW, Coates MI, Rubidge BS (2006) A lamprey from the devonian period of South Africa. Nature 443: 981–984

Grossmann M, Weintraub BD, Szkudlinski MW (1997) Novel insights into the molecular mechanisms of human thyrotropin action: structural, physiological, and therapeutic implications for the glycoprotein hormone family. Endocr Rev 18: 476–501

Gudermann T, Birnbaumer M, Birnbaumer L (1992) Evidence for dual coupling of the murine luteinizing hormone receptor to adenylyl cyclase and phosphoinositide breakdown and Ca^{2+} mobilization. Studies with the cloned murine luteinizing hormone receptor expressed in L cells. J Biol Chem 267: 4479–4488

Hauser F, Nothacker HP, Grimmelikhuijzen CJ (1997) Molecular cloning, genomic organization, and developmental regulation of a novel receptor from *Drosophila melanogaster* structurally related to members of the thyroid-stimulating hormone, follicle-stimulating hormone, luteinizing hormone/choriogonadotropin receptor family from mammals. J Biol Chem 272: 1002–1010

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Development 120 (Suppl): 125–133

Hoshii T, Takeo T, Nakagata N, Takeya M, Araki K, Yamamura K-I (2007) Lgr4 regulates the postnatal development and integrity of male reproductive tracts in mice. Biol Reprod 76: 303–313

Hsu SY, Liang SG, Hsueh AJ (1998) Characterization of two LGR genes homologous to gonadotropin and thyrotropin receptors with extracellular leucine-rich repeats and a G protein-coupled, seven-transmembrane region. Mol Endocrinol 12: 1830–1845

Janvier P (2006) Palaeontology: modern look for ancient lamprey. Nature 443: 921–924

Knudsen B, Miyamoto MM (2001) A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. Proc Natl Acad Sci USA 98: 14512–14517

Kudo M, Chen T, Nakabayashi K, Hsu SY, Hsueh AJ (2000) The nematode leucine-rich repeat-containing, G protein-coupled receptor (LGR) protein homologous to vertebrate gonadotropin and thyrotropin receptors is constitutively active in mammalian cells. Mol Endocrinol 14: 272–284

Kumar RS, Trant JM (2001) Piscine glycoprotein hormone (gonadotropin and thyrotropin) receptors: a review of recent developments. Comp Biochem Physiol B 129: 347–355

Lassmann T, Sonnhammer ELL (2005) Automatic assessment of alignment quality. Nucleic Acids Res 33: 7120–7128

Lei Y, Hagen GM, Smith SML, Barisas BG, Roess DA (2005) Chimeric GNRH-LH receptors and LH receptors lacking C-terminus palmitoylation sites do not localize to plasma membrane rafts. Biochem Bioph Res Co 337: 430–434

Loy TV, Vandersmissen HP, Hiel MBV, Poels J, Verlinden H, Badisco L, Vassart G, Broeck JV (2008) Comparative genomics of leucine-rich repeats containing G protein-coupled receptors and their ligands. Gen Comp Endocrinol 155: 14–21

Matsuo H, Baba Y, Nair RM, Arimura A, Schally AV (1971) Structure of the porcine LH- and FSH-releasing hormone. i. The proposed amino acid sequence. Biochem Bioph Res Co 43: 1334–1339

Mizutori Y, Chen C-R, McLachlan SM, Rapoport B (2008) The thyrotropin receptor hinge region is not simply a scaffold for the leucine-rich domain but contributes to ligand binding and signal transduction. Mol Endocrinol 22: 1171–1182

Moyle WR, Bahl OP, Marz L (1975) Role of carbohydrate of human chorionic gonadotropin in the mechanism of hormone action. J Biol Chem 250: 9163–9169

Moyle WR, Campbell RK, Myers RV, Bernard MP, Han Y, Wang X (1994) Co-evolution of ligand-receptor pairs. Nature 368: 251–255

Moyle WR, Lin W, Myers RV, Cao D, Kerrigan JE, Bernard MP (2005) Models of glycoprotein hormone receptor interaction. Endocrine 26: 189–205

Nakabayashi K, Matsumi H, Bhalla A, Bae J, Mosselman S, Hsu SY, Hsueh AJW (2002) Thyrostimulin, a heterodimer of two new

human glycoprotein hormone subunits, activates the thyroid-stimulating hormone receptor. J Clin Invest 109: 1445–1452

Nothacker HP, Grimmelikhuijzen CJ (1993) Molecular cloning of a novel, putative G protein-coupled receptor from sea anemones structurally related to members of the FSH, TSH, LH/CG receptor family from mammals. Biochem Biophys Res Commun 197: 1062–1069

Ohno S (1970) Evolution by Gene Duplication. Springer Verlag, New York

Scott DJ, Layfield S, Yan Y, Sudo S, Hsueh AJW, Tregear GW, Bathgate RAD (2006) Characterization of novel splice variants of LGR7 and LGR8 reveals that receptor signaling is mediated by their unique low density lipoprotein class a modules. Mol Endocrinol 20: 1880–1893

Seoighe C, Johnston CR, Shields DC (2003) Significantly different patterns of amino acid replacement after gene duplication as compared to after speciation. Mol Biol Evol 20: 484–490

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6: 715–722

Simoni M, Gromoll J, Nieschlag E (1997) The follicle-stimulating hormone receptor: biochemistry, molecular biology, physiology, and pathophysiology. Endocr Rev 18: 739–773

Sower SA, Moryiama S, Kasahara M, Takahashi A, Nozaki M, Uchida K, Dahlstrom JM, Kawauchi H (2006) Identification of sea lamprey GTHbeta-like cDNA and its evolutionary implications. Gen Comp Endocrinol 148: 22–32

Szkudlinski MW, Fremont V, Ronin C, Weintraub BD (2002) Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationships. Physiol Rev 82: 473–502

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Vischer HF, Bogerd J (2003a) Cloning and functional characterization of a gonadal luteinizing hormone receptor complementary DNA from the African catfish (*Clarias gariepinus*). Biol Reprod 68: 262–271

Vischer HF, Bogerd J (2003b) Cloning and functional characterization of a testicular TSH receptor cDNA from the African catfish (*Clarias gariepinus*). J Mol Endocrinol 30: 227–238

Wallace IM, O'Sullivan O, Higgins DG, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res 34: 1692–1699

You S, Kim H, Hsu CC, Halawani MEE, Foster DN (2000b) Three different turkey luteinizing hormone receptor (tLH-R) isoforms I: Characterization of alternatively spliced tLH-R isoforms and their regulated expression in diverse tissues. Biol Reprod 62: 108–116

You S, Kim H, Halawani MEE, Foster DN (2000a) Three different turkey luteinizing hormone receptor (tLH-R) isoforms II: Characterization of differentially regulated tLH-R messenger ribonucleic acid isoforms in the ovary. Biol Reprod 62: 117–124

(Received June 10, 2008 / Accepted September 14, 2008)