Endocrine, *vol. 26, no. 3, 267–276, April 2005* 0969–711X/05/26:267–276/$30.00

Conservation of the Heterodimeric Glycoprotein Hormone Subunit Family Proteins and the LGR Signaling System from Nematodes to Humans

Jae-Il Park, Jenia Semyonov, Chia Lin Chang,[1] and Sheau Yu Teddy Hsu

Sheau Yu T. Hsu, Division of Reproductive Biology, Department of Obstetrics and Gynecology, Stanford University School of Medicine, 300 Pasteur Drive, Room A344E, Stanford, CA 94305-5317; and [1]Department of Obstetrics and Gynecology, Chang-Gung Memorial Hospital, Tao-Yuan, Taiwan

Glycoprotein hormones, follicle-stimulating hormones (FSHs), luteinizing hormones (LHs), thyroid-stimulating hormones (TSHs), and chorionic gonadotropin (CG) are key endocrine hormones secreted from the pituitary gonadotrophs and thyrotrophs and the placenta in primates. These hormones, consisting of a common alpha subunit and a specific beta subunit, act through the FSH receptor (FSHR), the LH receptor (LHR), and the TSH receptor (TSHR) that are highly specific for their cognate hormones. These glycoprotein hormones are structurally and functionally conserved in various vertebrates and have been identified in most lineages of actinopterygians (bony fish) and sarcopterygians (tetrapods). Of interest, recent genomic studies showed that vertebrate glycoprotein hormone receptors belong to an ancient subfamily of G protein–coupled receptors (GPCRs) named as leucine-rich repeat-containing GPCRs (LGRs). These findings have prompted the hypothesis that there could be additional glycoprotein hormones in vertebrate genomes. Indeed, searches of vertebrate genomes have led to the identification of two novel glycoprotein hormone subunits, glycoprotein alpha 2 (GPA2) and glycoprotein beta 5 (GPB5), as well as their homologs in invertebrates. Subsequently, it was demonstrated that GPA2 and GPB5 form a heterodimeric hormone, thyrostimulin/OGH, capable of activating TSHR in vivo and the thyroid axis in transgenic mice. However, the exact role of this novel glycoprotein hormone and its homolog in invertebrates is not clear. To gain a better understanding of the physiological role of the novel glycoprotein hormone subunits and their evolution, it is imperative to carry out systematic studies of these genes in representative model species. In the present report, we summarize our findings based on studies of genomes of model organisms from sea anemones to humans. We found that GPA2 and GPB5 represent the ancient forms of glycoprotein hormone alpha and beta subunits, respectively, and that vertebrate and invertebrate glycoprotein hormone subunit proteins shared common ancestors that evolved during early metazoan evolution. It is important to note that glycoprotein hormone alpha and beta subunit proteins from invertebrates formed a heterodimer with structural functional characteristics similar to that of vertebrate glycoprotein hormones. Taken together, both glycoprotein hormone alpha and beta subunits evolved before the evolution of nematodes, arthropods, and vertebrates. Parallel expansion of the alpha and beta subunits and their receptors through gene duplication and subsequent subfunctionalization and neofunctionalization of the duplicated genes allowed the development of multiple tissue-specific endocrine systems in vertebrates.

Key Words: Glycoprotein hormone; gonadotropin; thyrotropin; GPCR; LGR; TSH; FSH; LH; hCG; GPA2; GPB5; thyrostimulin.

Introduction

Glycoprotein Hormones are Heterodimers Consisting of Two Cystine-Knot-Containing Proteins

Glycoprotein hormones, follicle-stimulating hormones (FSHs), luteinizing hormones (LHs), thyroid-stimulating hormones (TSHs), and choriogonadotropin (CG) are essential for gonadal and thyroid functions in vertebrates *(1–15)*. Gonadotropins regulate the growth and differentiation of the gonad, whereas TSH is essential for energy homeostasis *(3,4)*. Glycoprotein hormones are heterodimers consisting of two cystine-knot-containing proteins, a common alpha and a specific beta subunit *(16,17)*. The alpha subunit combines with four distinct beta subunits giving rise to four biologically active hormones in humans: FSH, LH, TSH, and CG. These hormones are highly conserved in organisms from primitive rayfin fish (*Chondrostei*) to humans in both primary sequences and functional characteristics *(18–20)*.

Received April 12, 2005; Accepted April 26, 2004.

Author to whom all correspondence and reprint requests should be addressed: Sheau Yu T. Hsu, Division of Reproductive Biology, Department of Obstetrics and Gynecology, Stanford University School of Medicine, 300 Pasteur Drive, Room A344E, Stanford, CA 94305-5317. E-mail: teddyhsu@stanford.edu

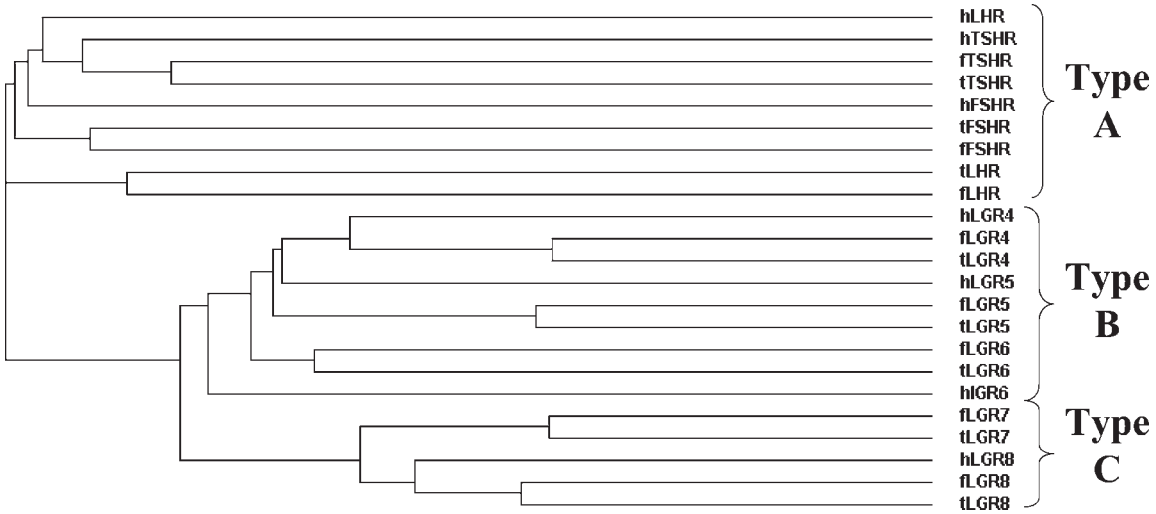

Fig. 1. Vertebrates share a similar inventory of LGRs. Phylogenetic relatedness of LGRs from human, pufferfish *Fugu rubripes,* and pufferfish *Tetraodon nigroviridis* were analyzed by the Clustal W method. The LGRs from both species can be divided into three distinct clusters. Type A LGR includes FSHR, LHR, and TSHR, whereas the LGR4, LGR5, and LGR6 orthologs form a separate cluster. The relaxin receptors, LGR7 and LGR8, belong to the third subgroup of LGRs. h: human, f: pufferfish *F. rubripes*, t: pufferfish *T. nigroviridis.*

They activate their respective receptors—the FSH receptor (FSHR), the LH receptor (LHR), and the TSH receptor (TSHR)—in target tissues *(16,18,19,21–25)*. The CG beta subunit of human choriogonadotropin (hCG), derived from recent duplications of the LH beta subunit gene, is unique to primates and important for pregnancy maintenance by activating the LH receptor *(26,27)*.

Studies on the crystal structure of the alpha and beta subunits of hCG have shown that glycoprotein hormone subunit polypeptides adopted a cystine-knot core structure shared by other signaling hormones including TGF-β, PDGF, NGF, and BMP antagonist family proteins *(2,28–32)*. Glycoprotein hormones represent a major part of the hypothalamus/pituitary/peripheral tissue (HPA) axis in the endocrine system of vertebrates, and were considered to be specific adaptations of vertebrates. However, recent molecular and bioinformatic studies revealed that various vertebrates and invertebrates contain novel orphan G protein–coupled receptors (GPCR) homologous to mammalian gonadotropin and thyrotropin receptors *(33–38)*. These receptors were named as leucine-rich repeat (LRR)–containing G protein–coupled receptors (LGRs) for the unique domain arrangement not found in other GPCRs. In addition to FSHR, LHR, and TSHR, most vertebrates contain five additional paralogous LGRs, named LGR4 to LGR8. All these receptors contain an extracellular domain with multiple LRRs followed by a GPCR transmembrane domain. LGRs can be divided into three subgroups based on sequence similarity (Fig. 1). The type A LGRs include the classic glycoprotein hormone receptors FSHR, LHR, and TSHR. In contrast, the type B LGR contains three orphan receptors, LGR4 to LGR6, whereas the type C LGR includes only two members, LGR7 (relaxin receptor 1) *(39–41)* and LGR8 (relaxin receptor 2 or INSL3 receptor) *(42–45)*. Based on studies of type A and type C LGRs, it has been shown that the ectodomain of these receptors functions as the ligand-binding domain that leads to the activation of the protein kinase A–dependent pathway in target tissues. Because all three types of LGRs were identified in insects, and type A LGR is present in most lineages of metazoa, it has been hypothesized that glycoprotein hormone subunit genes evolved early during evolution *(32)*. One of the great challenges in biology is to understand how complex functional adaptation originated and the underlying mechanisms. In this article, we address the origin of the vertebrate glycoprotein hormones and the coevolved type A LGRs.

Three Subtypes of LGRs Interact with Three Distinct Groups of Cystine-Knot-Containing Ligands

Studies of LGRs in model organisms indicated that these receptors are highly conserved from sea anemones to mammals, and that vertebrates share a similar inventory of LGRs (Fig. 1). Orthologs of human FSHR, LHR, TSHR, and LGR4–LGR8 were identified in teleosts (Fig. 1). Among the three subtypes of LGRs, type A LGR is the most widely distributed and has been identified in sea anemones, nematodes, mollusks, crustaceans, insects, and vertebrates *(36–38,40, 46,47)*. In contrast, type B and type C LGRs were identified in mollusks, insects, and vertebrates, but not in nematodes *(36,37,48–50)*. These data suggest that the three subtypes of LGRs could have evolved in the common ancestor of nematodes, arthropods, and vertebrates, and may interact with ligands with distinct properties. Therefore, the absence of type B and type C LGRs in nematodes could be due to gene loss during evolution. However, one cannot rule out the possibility that type B and type C LGRs evolved from

a type A LGR and experienced a greater sequence divergence after the separation of the nematodes and other metazoa. Functional analysis of LGRs showed that type A and type C LGRs from vertebrates interact with glycoprotein hormones and relaxin family peptides, respectively. Whereas the cognate ligands for type B LGRs in vertebrates are unknown, it was shown recently that the single type B LGR from *Drosophila melanogaster* mediates signaling by the tanning hormone bursicon *(46,47,51,52)*. Similar to glycoprotein hormones, bursicon was found to be a heterodimer consisting of two cystine-knot-containing proteins, burs and pburs *(46,47)*. Based on this evidence, we hypothesized that type B LGR in other invertebrates could be the cognate receptor for the burs and pburs homologs found throughout insects, crustaceans, and mollusks *(46)*.

Identification and Comparison of Glycoprotein Hormone Subunit Family Genes in Invertebrates

Analyses of type A LGRs from different metazoa indicated that type A LGRs are highly conserved, and the sequence conservation extends from the N-terminal cysteine-rich region to the transmembrane domain of these receptors *(35,36)*. This suggests that type A LGRs from invertebrates likely interact with a ligand consisting of two glycoprotein hormone subunit-like proteins.

Subsequent studies of the complete genome of yeast, *Caenorhabditis elegans, D. melanogaster*, and various invertebrates showed that there is one glycoprotein hormone beta subunit-like gene (GPB) in the genomes of nematode *C. elegans,* hookworm *Ancylostoma caninum*, and *D. melanogaster (32)*. In addition, vertebrates contain two additional glycoprotein hormone subunit family genes. The two novel human glycoprotein hormone subunit homologs share the conserved cysteine residue arrangement found in vertebrate glycoprotein hormone alpha subunit and beta subunit, respectively, and were named as glycoprotein alpha2 (GPA2) and glycoprotein beta5 (GPB5) *(32,53,54)*. Likewise, sequence alignment revealed that cystine-knot-forming cysteine residues in the invertebrate GPBs are in almost perfect alignment with their cysteine counterparts in vertebrate glycoprotein hormone beta subunits. In addition, human GPB5 shares the greatest sequence identity (37–42%) with gonadotropin beta subunits from Acipenseriformes (sturgeon) and invertebrate GPBs when compared to other beta subunits from mammals, suggesting that GPB5 and beta subunits from Actinopterygii fish retain the structural characteristics of the common ancestor of beta subunit family proteins and could represent the ancient form of glycoprotein hormone beta subunit family genes *(20)*. In addition, these earlier studies indicated that invertebrate GPB could interact with a glycoprotein hormone alpha subunit-like protein (GPA) to form a heterodimeric ligand for the single type A LGR found in various invertebrates.

Functional Characterization of Human GPA2 and GPB5

FSH, LH, and TSH are mainly expressed in the anterior pituitary to coordinate endocrine regulation in the hypothalamus/pituitary axis *(3–10)*. Unlike gonadotropin and thyrotropin subunit genes, the newly identified GPA2 and GPB5 were expressed not only in the pituitary but also in several peripheral tissues as well, thereby suggesting that these factors could regulate physiology in a manner distinct from that of known glycoprotein hormones *(32,53,54)*. Because the vertebrate GPB5 forms a separate cluster with invertebrate GPB, we hypothesized that GPB5 could represent part of a ligand of the type A LGRs *(32)*, but not type B or type C LGRs. Based on this understanding, we tested the bioactivity of GPA2 and GPB5 on type A LGRs. Consistent with this hypothesis, functional studies demonstrated that GPA2 and GPB5 formed a heterodimeric ligand, thyrostimulin, selective for the TSHR, but not LHR or FSHR in transfected cells *(53)*. In addition, injection of thyrostimulin in vivo led to phenotypes similar to that following TSH treatment *(53)*.

Consistent with the finding that the heterodimeric thyrostimulin activates TSHR, a parallel study showed that overexpression of thyrostimulin (or OGH) in transgenic mice leads to elevations of basal thyroid levels twofold of that found in wild-type littermates *(54)*. Because of an increase in their metabolic rates, transgenic mice weigh less than the controls despite increased food intake. In addition, transgenic mice with overexpression of thyrostimulin/OGH gain significantly less weight and body fat than controls when challenged with a high-fat diet. The activation of the thyroid axis in these transgenic mice also led to a reduction of blood glucose, insulin, cholesterol, and triglycerides *(54)*. Interestingly, unlike mice treated with exogenous thyroid hormones, mice with overexpression of thyrostimulin/OGH do not exhibit significantly elevated heart rates. Although the exact mechanism is not clear, it is possible that the TSH and thyrostimulin/OGH could exhibit subtle differences in the activation of TSHR. These studies showed that TSHR bioactivity could be modulated by a second ligand consisting of GPA2 and GPB5. Nonetheless, in a parallel study it was found that deletion of GPB5 in transgenic mice leads to no particular phenotype in body weight, response to high-fat diet, metabolic parameters, body composition, or insulin tolerance *(54)*.

TSH and the TSHR are key proteins in the control of thyroid function *(11–15)*. TSH synthesis and release is stimulated by hypothalamic TSH-releasing hormones and is regulated by thyroid hormones in a classic negative-feedback loop. The lack of signaling through the TSHR leads to thyroid hypoplasia, whereas overstimulation of the TSHR signaling leads to thyroid hyperplasia *(11–15)*. Because thyroid development is arrested in mice with targeted disruption of the common alpha subunit, it was believed that TSH is the sole ligand for TSHR. The finding that thyrostimulin/OGH

acts as a second ligand for TSHR and is capable of activating the thyroid axis in transgenic mice indicated that in vertebrates there are two TSHR ligands subserving different aspects of energy metabolism. However, the observation that GPB5 is dispensable for normal function in transgenic mice suggested that the newly identified thyrostimulin/OGH has a limited role in TSHR signaling in mice and the role of thyrostimulin/OGH in other species remains to be explored *(54)*.

Because GPA2 and GPB5 show a closer relationship to subunits from invertebrates, thyrostimulin/OGH could thus represent an ancient form of the TSHR ligand and, together with TSH, subserve part of the ligand–TSHR signaling pathway in modern vertebrates. However, whether thyrostimulin/OGH regulates a similar set of bioactivity in different vertebrates remains to be characterized.

Both Glycoprotein Hormone Alpha and Beta Subunit Genes Evolved Early During Metazoan Evolution

To gain a better understanding of the role of GPA2 and GPB5 in vertebrates and the evolution of glycoprotein hormones, we analyzed glycoprotein hormone subunit genes in model organisms. The search revealed that GPB-like genes could be traced to diverse invertebrates including fruitfly, mosquito, honeybee, sea hare, and nematode (Fig. 2A). As shown in Fig. 2B, alignment of the GPB from two species of nematodes, *C. elegans* and *C. briggsae*, indicated that the nematode GPB shared seven conserved cysteine residues with almost invariant spacing with the hCG beta subunit. In addition, it is evident that the invertebrate GPB forms a separate cluster with the GPB5 from vertebrates, further indicating that the GPB5 represents the ancient form of beta subunit family proteins (Fig. 2A). Interestingly, both zebrafish and puffer *Tetraodon nigroviridis* contain more than one copy of GPB5 (Fig. 2A). Thus, it is possible that in some species of vertebrate, the TSHR signaling system has evolved to interact with three distinct ligands: one TSH and two thyrostimulin/OGH homologs.

In addition, analysis of invertebrate genomes revealed the presence of GPA family proteins in various invertebrates including fruit fly, mosquito, and nematode *(60)*. These invertebrate GPAs showed a greater sequence similarity to the vertebrate GPA2 in vertebrates (Fig. 3A) indicating that, similar to GPB5, the GPA2 represents the ancient form of glycoprotein hormone alpha subunit family proteins. Alignment of the two nematode GPAs with the human common alpha subunit showed that although they exhibited less than 20% sequence identity, cysteine residues important for cystine-knot formation are conserved (Fig. 3B).

Furthermore, sequence analysis showed that nematode GPAs share great sequence similarity with the mature region of the BMP antagonist, gremlin, from diverse species. Hence, it is possible that these two cystine-knot-containing proteins share a common ancestor that formed homodimers or heterodimers with other cystine-knot-containing proteins. By forming a complex with the ancestral beta subunit, the heterodimeric glycoprotein hormone functions as a ligand for the type A LGR. On the other hand, the gremlin ancestor evolved to interact with a TGF-β family protein and regulate its signaling as an antagonist. This finding suggested that, during evolution, sequence divergence and association with different cystine-knot-containing partners lead to the generation of distinct hormones playing agonist or antagonist roles in cell–cell communication.

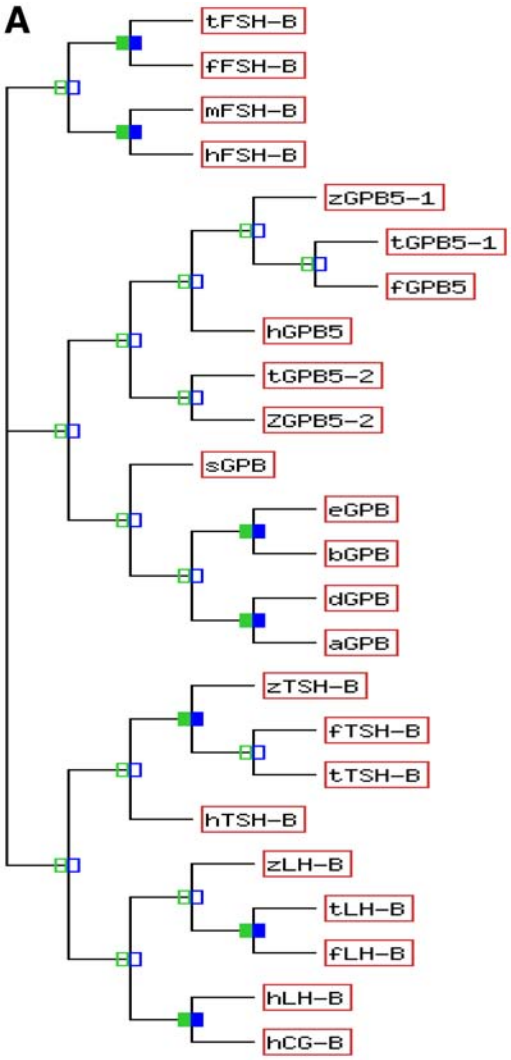

Fig. 2. Phylogenetic relatedness of glycoprotein hormone alpha subunit family proteins. (**A**) Phylogenetic analysis of glycoprotein hormone beta subunit family proteins from human, mouse, pufferfish *T. nigroviridis,* pufferfish *F. rubripes*, zebrafish *Danio rerio*, *D. melanogaster*, mosquito *Anopheles gambiae*, sea hare *Aplysia california*, nematode *C. elegans*, and nematode *C. briggsae*. h: human, m: mouse, t: pufferfish *T. nigroviridis,* f*:* pufferfish *F. rubripes,* z: zebrafish *D. rerio,* d: *D. melanogaster*, a: *A. gambiae,* s: sea hare *A. california;* e: nematode *C. elegans*, b: nematode *C. briggsae*.

Fig. 2. (**B**) Sequence alignment of the hCG beta subunit and GPB from nematodes, *C. elegans,* and *C. briggsae*. The structure-determining cysteine residues are highlighted. The cystine-knot-forming cysteines are indicated by asterisks. h: human, Ce: nematode *C. elegans*, Cb: nematode *C. briggsae*.

Invertebrate GPA and GPB Proteins Share Similar Structural and Functional Characteristics with Vertebrate Glycoprotein Hormones

To study whether the newly identified GPA and GPB from diverse invertebrates share the structural characteristics of vertebrate glycoprotein hormones, we generated the tertiary structure model of the putative GPA/GPB heterodimer from the nematode *C. briggsae* by comparative protein modeling using the crystal structures of hCG (Protein Data Bank accession code 1HRP) as the template *(55)*. As shown in Fig. 4A, the putative tertiary structure of nematode GPA/GPB heterodimer closely resembles that of the hCG and other vertebrate glycoprotein hormones along the entire length. Like hCG, the GPA and GPB of nematodes adopt a structure with two beta-hairpin loops below the cystine knot and a long loop above. The two subunits appear to interact with each other by similar interfaces found in the common alpha and CG beta subunits in vertebrates (Fig. 4A). The major variations between hCG and the nematode GPA/GPB heterodimer lie in the C-terminus of the beta subunit (Fig. 2B). As a result, GPB from nematodes and other invertebrates lacks the long C-terminal strand and the disulfide bridge linking this C-terminal strand to the second hairpin loop found in hCG beta subunit *(31,32)*. Surprisingly, although the nematode GPB lacks the seatbelt motif known to be important for heterodimer formation, the nematode GPA/GPB heterodimer appears to interact closely in a manner similar to vertebrate glycoprotein hormones (Fig. 4A, oblong area). Furthermore, the model revealed that a bulging motif on the hCG (Fig. 4A, circle area), represented by the N-terminal sequences of the CG beta subunit, was represented by the nematode GPA sequences in the nematode GPA/GPB heterodimer. These results indicated that the nematode GPA/GPB heterodimer shares a similar tertiary structure with that of hCG even though the two molecules have a very limited sequence similarity.

Furthermore, we modeled the ectodomain of nematode LGRs using the human FSH receptor ectodomain crystal structure (Fig. 4B). Analysis of the simulated model indicated that the ectodomain of nematode LGR adopts a structure similar to that of human FSHR where each ectodomain contains 10 leucine-rich repeats with the beta sheet of each LRR facing the ligand-binding interface. Like that of FSHR, LRRs at the middle of the ectodomain of nematode LGR have an extended beta sheet compared to those at the N- and C-terminus of the ectodomain. The major difference is that the nematode LGR has a truncated first LRR located at the N-terminus of the ectodomain. In addition to comparative modeling, functional analyses showed that the GPA/GPB heterodimer from *D. melanogaster* is the cognate ligand for the single type A LGR in *D. melanogaster*.

Expansion of Glycoprotein Hormone Subunits and the LGR Signaling Systems During Evolution

Recent studies of GPCR genes in various model organisms have shown that the number of GPCRs varied greatly in different lineages due to expansion of select GPCRs important for lineage-specific adaptation *(56–58)*. Because the ligands of GPCRs diverged significantly during evolution, it was difficult to assign an orthologous relationship among homologous GPCRs from invertebrate and vertebrates. Likewise, analysis of LGRs clearly showed that all three subgroups of LGRs expanded in the vertebrate and there is no definitive orthologous relationship between the vertebrate type A LGRs and the invertebrate type A LGRs. Although most invertebrates contain only one type A LGR, most vertebrates studied have at least three type A LGR paralogs. The three glycoprotein hormone receptors would have evolved from sequential duplication of an ancestral type A LGR that regulated select physiological functions in the common ancestral metazoa. Although the function of

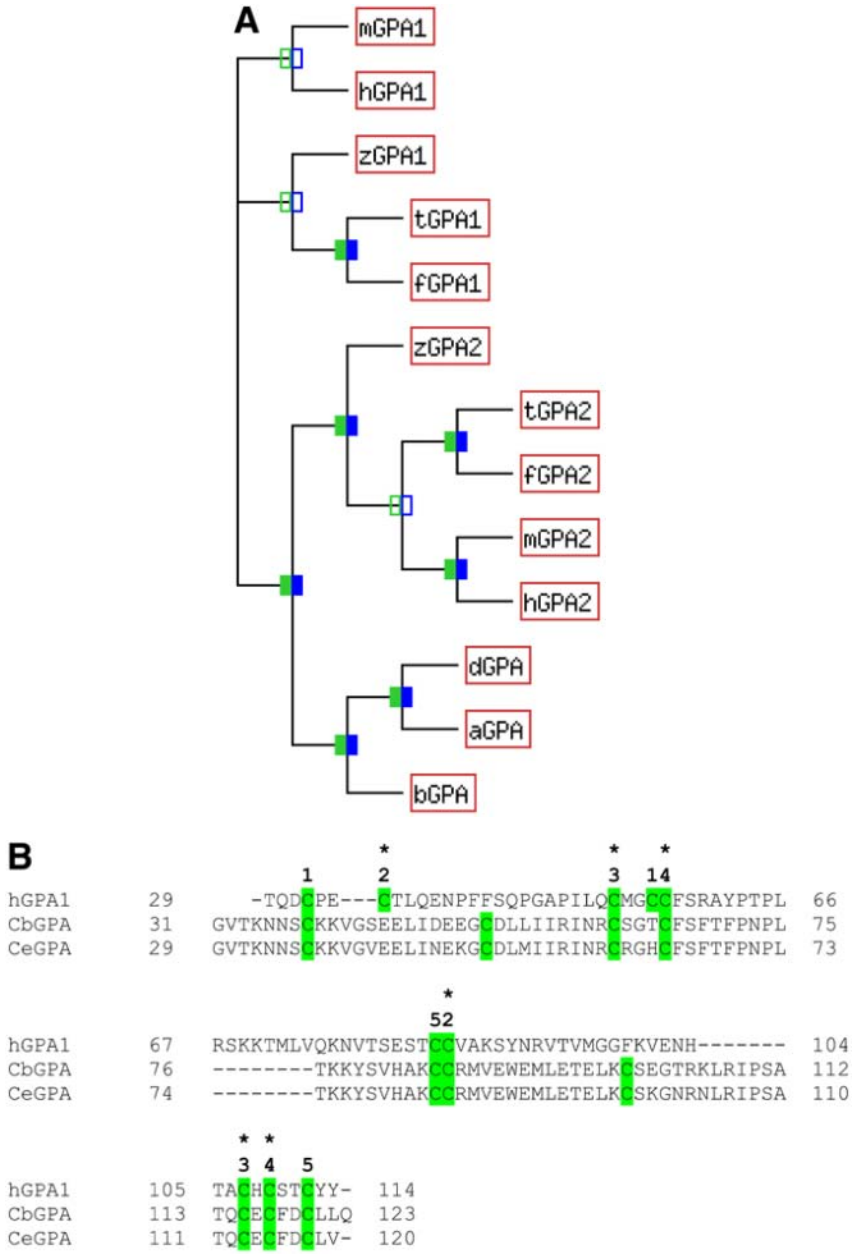

Fig. 3. Phylogenetic relatedness of glycoprotein hormone alpha subunit family proteins. (**A**) Phylogenetic analysis of glycoprotein hormone alpha subunit family proteins from human, mouse, pufferfish *T. nigroviridis,* pufferfish *F. rubripes*, zebrafish *D. rerio*, *D. melanogaster*, mosquito *A. gambiae*, and nematode *C. briggsae*. h: human, m: mouse, t: pufferfish *T. nigroviridis,* f*:* pufferfish *F. rubripes,* z: zebrafish *D. rerio,* d: *D. melanogaster*, a: *A. gambiae*, b: nematode *C. briggsae*. (**B**) Sequence alignment of the common glycoprotein hormone alpha subunit and GPA from nematodes, *C. elegans,* and *C. briggsae*. Structure-determining cysteine residues are highlighted. The cystine-knot-forming cysteines are indicated by asterisks. h: human, Ce: nematode *C. elegans*, Cb: nematode *C. briggsae*.

GPA/GPB heterodimer in invertebrates remains to be explored, the matching of this ligand with the type A LGR from *D. melanogaster* clearly demonstrated that the glycoprotein hormone receptor signaling has an ancient origin (*60*).

Similar to the evolution of receptors, modification of regulatory and structural elements of the duplicated ligand gene in a chordate ancestor through natural selection led to the divergence and increased specificity of duplicated glycoprotein hormone toward coevolved type A LGRs. A model of how the vertebrate glycoprotein hormone (GPA/GPB heterodimer) endocrine system could have emerged in relatively few steps by gene duplication is presented in Fig. 5. In this model, semi-parallel duplications of ancestral alpha and beta subunit genes led to the generation of multiple heterodimers. These changes resulted in new signaling connections. Through neofunctionalization and subfunctionalization of the ligand–receptor pairs *(59)*, the preexisting regulatory systems underwent divergence, leading to the diverged expression of duplicated paralogous glycoprotein hormones and their receptors in select tissues.

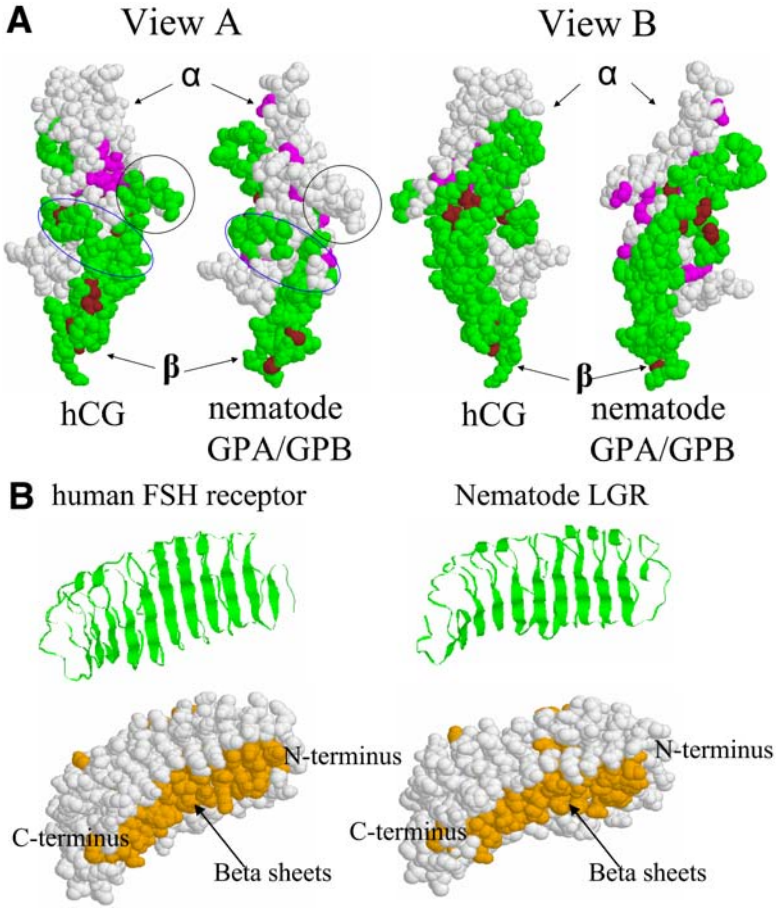

Fig. 4. Conservation of structures of the heterodimeric glycoprotein hormones and the type A LGR ectodomain from nematodes to humans. (**A**) Comparison of the tertiary structure of nematode GPA/GPB heterodimer with hCG. The GPA/GPB heterodimer model was generated based on the hCG crystal structure template. Views from two different angles (left, view A with the alpha subunit in the foreground; right, view B with the beta subunit in the foreground) are presented. The alpha subunits are indicated in white, whereas the beta subunit is gray. Cysteine residues that are important for the formation of structure-determining cystine bonds in the alpha and beta subunits are indicated by dark shading. The region formed by the C-terminus of the hCG beta subunit is circled by oblong elipse for comparison with the corresponding region in the nematode GPA/GPB heterodimer. Although the nematode GPB lacks the C-terminus for forming a seat-belt structure, the model predicted that the GPA and GPB from nematodes interact closely in the corresponding region. In addition, the N-terminus of the CG beta is indicated by a circle. This region is formed by the beta subunit residues in hCG, whereas in the corresponding region of nematode GPA/GPB heterodimer it was represented by residues from the nematode GPA. (**B**) Comparison of the tertiary structure of the ectodomain of nematode LGR and human FSHR. Upper panel, a ribbon model for the nematode LGR ectodomain is shown to the right of the FSHR ectodomain. Lower panel, a space-fill model of the nematode LGR ectodomain is shown to the right of the FSHR ectodomain. Beta sheets within each of the 10 LRRs in the receptor ectodomain are indicated by shading.

This model is best explained by the evolution of CGs from LH in the primate ancestor. The duplicated LH beta subunits diverged in their expression pattern, leading to the generation of hCG exhibiting a placenta-specific expression pattern and functional characteristics. Likewise, following parallel duplication of the ancestral alpha and beta subunit genes, TSH and thyrostimulin/OGH coevolved with one of the duplicated type A LGRs that eventually evolved as the TSHR. During evolution, subfunctionalization and neofunctionalization of these duplicated proteins evolved to regulate a subset of TSHR signaling *(59)*.

In addition to ancient gene duplication and the subsequent derivation of select specificity, vertebrate lineage-specific innovations of glycoprotein hormone subunit genes are also evident. For example, although most tetrapods have one LH, primates and equines have independently evolved CG and PMSG, respectively. It has been hypothesized that the evolution of hCG beta subunits with a long half-life could be a result of read-through mutation of the ancestral LH beta subunit gene. Likewise, some teleosts retain more than one copy of GPB5 (Fig. 2A) derived from lineage-specific tandem gene duplication or a genome duplication event.

Conclusions and Future Direction

Glycoprotein hormones coordinate growth, metabolism, and reproduction in all vertebrates. This system appears highly conserved across vertebrates including actinopterygians and

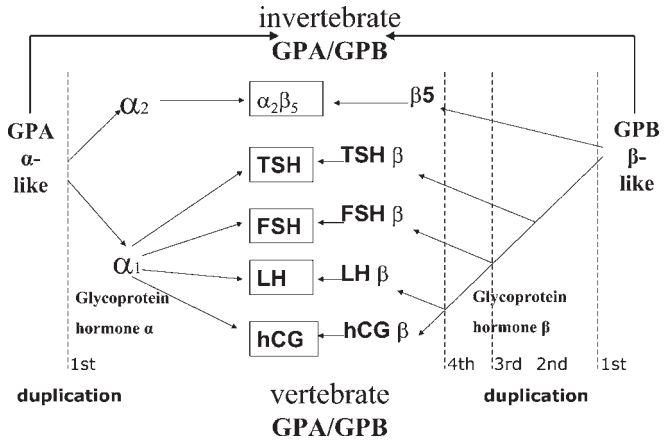

Fig. 5. Hypothetical model of the expansion of glycoprotein hormone subunit genes through gene duplication. Both glycoprotein hormone alpha and beta subunits evolved from ancestral genes that were present prior to the evolution of nematodes and vertebrates. Phylogenetic analyses of subunit genes from vertebrate and invertebrates indicated that the common alpha subunit (α_1) and GPA2 (α_2) in vertebrates derived from one gene duplication during early evolution of vertebrates. In contrast, the vertebrate glycoprotein hormone beta subunits evolved from four sequential gene duplications following the separation of chordates from invertebrates. On the other hand, only single glycoprotein hormone alpha and beta subunit genes were present in invertebrates investigated.

sarcopterygians. Studies of genome sequences revealed that glycoprotein hormones and type A LGRs have an ancient origin dating before the divergence of nematodes and vertebrates one billion years ago, and that the glycoprotein hormone–LGR signaling system expanded to regulate tissue-specific signaling in vertebrates. This finding is consistent with the evolutionary trend that derived genes regulate specialized functions through selection and random mutation.

Sequence analysis indicated that gonadotropin and thyrotropin diverged greatly from the GPA2 and GPB5 paralogs as well as the invertebrate homologs. The major divergence of these paralogs is the difference in the number and position of the structure-determining cysteine residues. For example, 2 of the 10 conserved cysteines found in the common alpha subunit and GPA2 are positioned differently. On the other hand, invertebrate GPB contains 7 cysteines whereas the hCG beta subunit has 12 cysteines. Owing to the absence of multiple cysteines and the C-terminal motifs in GPB5 and invertebrate GPB, it was thought that heterodimers consisting of these polypeptides would be depleted of the "seat-belt" wraps found in vertebrate glycoprotein hormones. However, structural changes to ancestral homologs of glycoprotein hormone subunit proteins appeared to be minimal as indicated by the structure model of the nematode GPA/GPB heterodimer (Fig. 4A). The interaction between GPA and GPB in invertebrates is snug and the two subunits burrow together in a manner similar to hCG.

Vertebrate features that are absent from invertebrates represent adaptation and innovations made after the divergence of these lineages. In mammals, in addition to TSH, the newly identified thyrostimulin/OGH is capable of regulating the thyroid axis and the production of other hormones by target tissues. That additional gonadotropin and thyrotropin paralogs are retained in the genome indicated that these ligands acquired distinct functions. In contrast, other duplicated genes such as TSH and thyrostimulin/OGH underwent subfunctionalization and each subserves select physiological processes *(54)*. This information, together with future studies of the bioactivity of thyrostimulin/OGH in different species, could aid in the understanding of the physiological role of thyrostimulin/OGH orthologs in different vertebrates.

The conservation of endocrine genes in vertebrates and invertebrates supports a model in which gene duplication and subsequent divergence in structural and regulatory elements led to the expanded glycoprotein hormone endocrine system in vertebrates. Unlike vertebrates, the invertebrate glycoprotein hormone system appeared to remain monophyletic and only one pair of ligands and receptors are present throughout nematodes and the insects. Similar to the ligand for type B LGR in insects, the invertebrate GPA/GPB heterodimer could function as a neuroendocrine factor regulating a physiological function mediated by the single type A LGR *(60)*. Future studies on the exact bioactivity of the invertebrate glycoprotein hormones and type A LGRs could reveal how the invertebrate neuroendocrine system evolves into the HPA axis in vertebrates.

Materials and Methods

Identification of Novel Glycoprotein Hormone Subunit Family Genes from Vertebrates and Invertebrates

To identify glycoprotein hormone subunit family genes, known glycoprotein hormone subunits were used as the query to search for homologous sequences in GenBank using the BLAST server at the National Center for Biotechnology Information (NIH, Bethesda, MD). To identify all potential paralogous genes in different species, the statistical significance threshold for reporting matches against database sequences was set at a less stringent level. Each recognized protein or DNA fragment was analyzed against the nonredundant database of GenBank to verify gene identity, followed by manual searches of invariably spaced cysteine residues important for cystine bond formation in known glycoprotein hormone subunits.

Sequence Alignment and Phylogenetic Analysis

Alignment and phylogenetic analyses of glycoprotein hormone subunit family proteins was carried out by pair-BLAST and Clustal W. The BLOCK Maker program (http://blocks.fhcrc.org) and the EMBL-EBI European Bioinformatics Institute server (http://www.ebi.ac.uk/clustalw) were used to align and generate the highly conserved blocks of aligned

sequences from different species. The phylogenetic relationship among glycoprotein hormone subunit family proteins was constructed by the neighboring-joining method from the Block alignments using a routine in ClustalW.

Structural Modeling

Comparative protein modeling was performed with the SWISS-MODEL server (http://swissmodel.expasy.org//SWISS-MODEL.html) using experimentally determined structures for human CG as the template (Protein Data Bank under the accession code 1HRP). DeepView-Swiss-Pdb Viewer and Protein Explorer (http://molvis.sdsc.edu/protexpl) were used to visualize the three-dimensional structure as well as to conduct structure comparisons of different family proteins.

Acknowledgments

We thank Caren Spencer for editorial assistance, and Dr. Aaron J. W. Hsueh for comments on the manuscript. Chia Lin Chang is supported by a Dean's Fellowship from the Chang Gung Memorial Hospital, Department of Obstetrics and Gynecology, Tao-Yuan, Taiwan. The work was partly supported by the Collaborative Research Initiative of the Specialized Cooperative Centers Program in Reproduction Research at Stanford University through cooperative agreement V54HD31398.

References

1. Hearn, M. T. and Gomme, P. T. (2000). *J. Mol. Recognit.* **13,** 223–278.
2. Isaacs, N. W. (1995). *Curr. Opin. Struct. Biol.* **5,** 391–395.
3. Kol, S. and Adashi, E. Y. (1995). *Curr. Opin. Obstet. Gynecol.* **7,** 209–213.
4. Jameson, J. L. (1996). *Mol. Cell. Endocrinol.* **125,** 143–149.
5. Achermann, J. C. and Jameson, J. L. (1999). *Mol. Endocrinol.* **13,** 812–818.
6. Hsueh, A. J., Eisenhauer, K., Chun, S. Y., Hsu, S. Y., and Billig, H. (1996). *Recent Prog. Horm. Res.* **51,** 433–455.
7. Themmen, A. P. N. and Huhtaniemi, I. T. (2000). *Endocr. Rev.* **21,** 551–583.
8. Puett, D., Bhowmick, N., Fernandez, L. M., Huang, J., Wu, C., and Narayan, P. (1996). *Mol. Cell. Endocrinol.* **125,** 55–64.
9. Huhtaniemi, I. (2000). *Horm. Res.* **53,** 9–16.
10. Hai, M. V., De Roux, N., Ghinea, N., et al. (1999). *Ann. Endocrinol. (Paris)* **60,** 89–92.
11. Brown, R. S. (2004). *Endocrinology* **145,** 4058–4061.
12. Vassart, G., Pardo, L., and Costagliola, S. (2004). *Trends Biochem. Sci.* **29,** 119–126.
13. Milgrom, E. (2003). *Bull. Acad. Natl. Med.* **187,** 671–680; discussion 680–682.
14. Rodien, P., Ho, S. C., Vlaeminck, V., Vassart, G., and Costagliola, S. (2003). *Ann. Endocrinol. (Paris)* **64,** 12–16.
15. Wonerow, P., Neumann, S., Gudermann, T., and Paschke, R. (2001). *J. Mol. Med.* **79,** 707–721.
16. Boime, I. and Ben-Menahem, D. (1999). *Recent Prog. Horm. Res.* **54,** 271–288.
17. Corless, C. L., Matzuk, M. M., Ramabhadran, T. V., Krichevsky, A., and Boime, I. (1987). *J. Cell. Biol.* **104,** 1173–1181.
18. So, W. K., Kwok, H. F., and Ge, W. (2005). *Biol. Reprod.* PMID: 15728794.
19. Kwok, H. F., So, W. K., Wang, Y., and Ge, W. (2005). *Biol. Reprod.* PMID: 15728795.
20. Querat, B., Sellouk, A., and Salmon, C. (2000). *Biol. Reprod.* **63,** 222–228.
21. Laan, M., Richmond, H., He, C., and Campbell, R. K. (2002). *Gen. Comp. Endocrinol.* **125,** 349–364.
22. Ryan, R. J., Charlesworth, M. C., McCormick, D. J., Milius, R. P., and Keutmann, H. T. (1988). *FASEB J.* **2,** 2661–2669.
23. Keutmann, H. T. (1992). *Mol. Cell Endocrinol.* **86,** C1–C6.
24. Segaloff, D. L., Sprengel, R., Nikolics, K., and Ascoli, M. (1990). *Recent Prog. Horm. Res.* **46,** 261–301.
25. Milgrom, E., de Roux, N., Ghinea, N., et al. (1997). *Horm. Res.* **48,** 33–37.
26. Lustbader, J. W., Lobel, L., Wu, H., and Elliott, M. M. (1998). *Recent Prog. Horm. Res.* **53,** 395–424.
27. Zeleznik, A. J. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 11002–11007.
28. Kaivo-Oja, N., Bondestam, J., Kamarainen, M., et al. (2003). *J. Clin. Endocrinol. Metab.* **88,** 755–762.
29. Vitt, U. A., Hsu, S. Y., and Hsueh, A. J. (2001). *Mol. Endocrinol.* **15,** 681–694.
30. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., and Hendrickson, W. A. (1994). *Structure* **2,** 545–558.
31. Lapthorn, A. J., Harris, D. C., Littlejohn, A., et al. (1994). *Nature* **369,** 455–461.
32. Hsu, S. Y., Nakabayashi, K., and Bhalla, A. (2002). *Mol. Endocrinol.* **16,** 1538–1551.
33. Nothacker, H. P. and Grimmelikhuijzen, C. J. (1993). *Biochem. Biophys. Res. Commun.* **197,** 1062–1069.
34. Eriksen, K. K., Hauser, F., Schiott, M., Pedersen, K. M., Sondergaard, L., and Grimmelikhuijzen, C. J. (2000). *Genome Res.* **10,** 924–938.
35. Nishi, S., Hsu, S. Y., Zell, K., and Hsueh, A. J. (2000). *Endocrinology* **141,** 4081–4090.
36. Hsu, S. Y., Kudo, M., Chen, T., et al. (2000). *Mol. Endocrinol.* **14,** 1257–1271.
37. Hsu, S. Y., Liang, S. G., and Hsueh, A. J. (1998). *Mol. Endocrinol.* **12,** 1830–1845.
38. Tensen, C. P., Van Kesteren, E. R., Planta, R. J., et al. (1994). *Proc. Natl. Acad. Sci. USA* **91,** 4816–4820.
39. Krajnc-Franken, M. A., van Disseldorp, A. J., Koenders, J. E., Mosselman, S., van Duin, M., and Gossen, J. A. (2004). *Mol. Cell Biol.* **24,** 687–696.
40. Hsu, S. Y. (2003). *Trends Endocrinol. Metab.* **14,** 303–309.
41. Hsu, S. Y., Nakabayashi, K., Nishi, S., et al. (2002). *Science* **295,** 671–674.
42. Kawamura, K., Kumagai, J., Sudo, S., et al. (2004). *Proc. Natl. Acad. Sci. USA* **101,** 7323–7328.
43. Kamat, A. A., Feng, S., Bogatcheva, N. V., Truong, A., Bishop, C. E., and Agoulnik, A. I. (2004). *Endocrinology* **145,** 4712–4720.
44. Bogatcheva, N. V., Truong, A., Feng, S., Engel, W., Adham, I. M., and Agoulnik, A. I. (2003). *Mol. Endocrinol.* **17,** 2639–2646.
45. Sudo, S., Kumagai, J., Nishi, S., et al. (2002). *J. Biol. Chem.* **278,** 7855–7862.
46. Luo, C. W., Dewey, E. M., Sudo, S., et al. (2005). *Proc. Natl. Acad. Sci. USA* **102,** 2820–2825.
47. Mendive, F. M., Loy, T. V., Claeysen, S., et al. (2005). *FEBS Lett.* **579,** 2171–2176.
48. Herpin, A., Badariotti, F., Rodet, F., and Favrel, P. (2004). *Biochim. Biophys. Acta* **1680,** 137–144.
49. Nishi, S., Hsu, S. Y., Zell, K., and Hsueh, A. J. (2000). *Endocrinology* **141,** 4081–4090.
50. Eriksen, K. K., Hauser, F., Schiott, M., Pedersen, K. M., Sondergaard, L., and Grimmelikhuijzen, C. J. (2000). *Genome Res.* **10,** 924–938.
51. Dewey, E. M., McNabb, S. L., Ewer, J., et al. (2004). *Curr. Biol.* **14,** 1208–1213.

52. Honegger, H. W., Dewey, E. M., and Kostron, B. (2004). *Acta Biol. Hung.* **55,** 91–102.
53. Nakabayashi, K., Matsumi, H., Bhalla, A., et al. (2002). *J. Clin. Invest.* **109,** 1445–1452.
54. Macdonald, L. E., Wortley, K. E., Gowen, L. C., et al. (2005). *Proc. Natl. Acad. Sci. USA* **102,** 2496–2501.
55. Erbel, P. J., Karimi-Nejad, Y., De Beer, T., Boelens, R., Kamerling, J. P., and Vliegenthart, J. F. (1999). *Eur. J. Biochem.* **260,** 490–498.
56. Hewes, R. S. and Taghert, P. H. (2001). *Genome Res.* **11,** 1126–1142.
57. Takeda, S., Kadowaki, S., Haga, T., Takaesu, H., and Mitaku, S. (2002). *FEBS Lett.* **520,** 97–101.
58. Zhang, X. and Firestein, S. (2002). *Nat. Neurosci.* **5,** 124–133.
59. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). *Genetics* **151,** 1531–1545.
60. Sudo, S., Kuwabara, Y., Park, J. L., Hsu, S. Y., and Hsueh, A. J. (2005). *Endocrinology,* in press.