An evolutionary comparison of leucine-rich repeat containing G protein-coupled receptors reveals a novel LGR subtype

Matthias B. Van Hiel a,b,1, Hans Peter Vandersmissen a,1, Tom Van Loy a, Jozef Vandenbergbroek a,*

a Animal Physiology and Neurobiology, Zoological Institute of the Katholieke Universiteit Leuven, Naamsestraat 59, P.O. Box 02465, B-3000 Leuven, Belgium
b Department of Genetics, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, 30 Flemington Road, Parkville, Melbourne, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 4 September 2011
Received in revised form 3 November 2011
Accepted 3 November 2011
Available online xxx

Keywords:
Leucine-rich repeat containing G protein-coupled receptor
Glycoprotein hormone receptor
Evolution

ABSTRACT

Leucine-rich repeat containing G protein-coupled receptors or LGRs are receptors with important functions in development and reproduction. Belonging to this evolutionarily conserved group of receptors are the well-studied glycoprotein hormone receptors and relaxin receptors in mammals, as well as the bursicon receptor, which triggers cuticle hardening and tanning in freshly eclosed insects. In this study, the numerous LGR sequences in different animal phyla are analyzed and compared. Based on these data a phylogenetic tree was generated. This information sheds new light on structural and evolutionary aspects regarding this receptor group. Apart from vertebrates and insects, LGRs are also present in early chordates (Urochordata, Cephalochordata and Hyperoartia) and other arthropods (Arachnida and Branchiopoda) as well as in mollusca, Echinodermata, Hemichordata, Nematoda, and even in ancient animal life forms, such as Cnidaria and Placozoa. Three distinct types of LGR exist, distinguishable by their number of leucine-rich repeats (LRRs), their type-specific hinge region and the presence or absence of an LDLα motif. Type C LGRs containing only one LDLα (C1 subtype) appear to be present in nearly all animal phyla. We here describe a second subtype, C2, containing multiple LDLα motifs, which was discovered in echinoderms, mollusks and in one insect species (Pediculus humanis corporis). In addition, eight putative LGRs can be predicted from the genome data of the placozoan species Trichoplax adhaerens. They may represent an ancient form of the LGRs, however, more genomic data will be required to confirm this hypothesis.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

G protein-coupled receptors (GPCRs) containing an N-terminal domain with leucine-rich repeats (LRRs), such as the mammalian glycoprotein hormone receptors, have long been thought to be restricted to vertebrates. However, based on molecular cloning efforts and genome analyses, the ancient origin of the leucine-rich repeat containing GPCRs (LGRs) was revealed. Three distinct types of LGRs have been defined based on their structural characteristics (type A, B and C). The differences between these three LGR types mainly lie in the structure of their ectodomain, while the serpentine region is relatively well preserved between the three types. The three LGR types differ in the number of LRR motifs, the absence or presence of an LDLα motif (low density lipoprotein-receptor-like cysteine-rich motif) and the type-specific hinge region (Fig. 1). An LRR motif consists of approximately 24 amino acids (AAs) and has a consensus sequence LxxLxxLxxLxxNxxLxxLxxPxxxPxx in which the leucines (L) can be readily replaced by an isoleucine (I) or a valine (V) and sometimes even with the weakly similar phenylalanine (F) or methionine (M) (Suppl. File 1A). Generally, type B LGRs have about twice the number of LRRs compared to the other two types. An exclusive feature of the type C LGRs is the presence of an LDLα motif in the ectodomain.

The sequence of the hinge region appears to be specific for every type of LGR. It is the sequence between the most C-terminal leucine-rich repeat and the start of the transmembrane domain. It functions, as the name implies, as a hinge between the anchored transmembrane domain and the extracellular horseshoe-shaped ectodomain. Type A LGRs include the mammalian glycoprotein hormone receptors FSHR, LH/CGR and TSHR and the fruit fly LGR1. The type B LGRs contain the mammalian LGR4, 5 and 6, and the fruit fly LGR2 or bursicon receptor. The third type of LGRs (type C) contain the mammalian LGR7 and LGR8, also called RXFP1 and RXFP2 respectively, and the fruit fly LGR3 and LGR4 [67]. Some LGRs have already been

Abbreviations: FSH(R), Follicle stimulating hormone (receptor); GPCR, G protein-coupled receptor; ILP, Insulin-like peptide; LDLα, low density lipoprotein receptor-like motif; LGR, Leucine-rich repeat containing GPCR; LH/CG(R), Luteinizing hormone/choriogonadotropin (receptor); LRR, Leucine-rich repeat; TSHR, Thyroid stimulating hormone (receptor).

* Corresponding author. Tel.: +32 16324260; fax: +32 16323902.
E-mail addresses: mvh@unimelb.edu.au (M.B. Van Hiel), Hans.Peter.Vandersmissen@bio.kuleuven.be (H.P. Vandersmissen), Tom.Van.Loy@bio.kuleuven.be (T. Van Loy), Jozef.VandenBroeck@bio.kuleuven.be (J. Vandenbergbroek).

1 These authors contributed equally.

0196-9781/5 – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.peptides.2011.11.004

functionally characterized, such as the mammalian relaxin receptors [27,37] and glycoprotein hormone receptors, and fruit fly LGR2 [47].

Using genome databases available at the NCBI platform (National Center for Biotechnology Information) (www.ncbi.nlm.nih.gov) and Ensembl (wwwensembl.org) and a variety of genome project databases from individual species, such as the Californian purple sea urchin (www.spbase.org) or the water flea (wfleabase.org), we predicted new LGR sequences in an attempt to generate a more comprehensive phylogenetic tree containing all presently known and predicted LGRs. This comparative analysis will contribute to a better understanding of the structural diversity and evolution of LGRs in Metazoa.

2. Methods

Sequences were found either by keyword searches or by using the (PSI-)BLAST algorithm against all known LGRs. Protein sequences were checked for appropriate domains using the internet software program SMART [39,60]. Possible incorrectly in silico predicted sequences were manually checked at the genomic DNA level and, where possible, repredicted using softberry (http://linux1.softberry.com/berry.phtml) FGENESH+. Sequences were subsequently manually corrected using the “six-frame translation” software of 2bio.net (http://molbiol.ru/index.php?langid=en). By searching for protein domains and through comparison with the most closely related LGRs, aligned using ClustalX2 [25,38], exons were identified. New manually predicted sequences were added in a supplementary file (Suppl. File 1B). To check and further complement previous search results, novel sequences were used as additional queries.

Phylogenetic trees were generated using MUSCLE alignments [18] and PhyML [22] (WAG Substitution Model, 500 bootstraps) at the Phylogeny.fr server [15]. Trees were rendered with the online TreeDyn application [11] and exported to MEGA [64]. Only complete or nearly complete sequences were used to avoid incorrect clustering as much as possible.

3. Results

3.1. Three distinctive types of LGRs exist

As with previously generated trees [68], three distinctly organized receptor types can be identified from our phylogenetic analysis. For a comprehensive overview, a phylogenetic tree generated based on a representative selection of LGR sequences from all phyla known to contain these receptors can be viewed in Suppl. File 2. A simplified version of this tree can be found in Fig. 2. When viewing this tree it can be concluded that, in addition to the vertebrates, arthropods (the water flea Daphnia arenata and Drosophila melanogaster) and echinoderms (the sea urchin Strongylocentrotus purpuratus) also possess one or more members of each LGR type. The most basal animal phylum in which LGRs have been described is Cnidaria. In Hydra magnipapillata and Nematostella vectensis, members of this phylum, one sequence of the type A and type B LGR groups have been identified. In molluskan species, up to now, a single type B (in the bivalve Crassostrea gigas) [23] and a single type C LGR (in the snail Lymnaea stagnalis [65]) were uncovered.

3.2. Type A LGRs

The LGRs belonging to type A typically have 7–9 LRRs in their ectodomain. Furthermore, they have a long hinge region, containing the consensus sequence LxxAXTYPxHCCAF at the beginning of the hinge and the sequence VxCXPxPDAFNPEDIMGYxFLRV at the end of the hinge (Suppl. File 3). In addition to the two cysteines in each of the consensus sequences, two more cysteines are present.
in the hinge region. Likely, those six cysteines form three disulfide bridges (S–S) stabilizing the entire structure. The hinge regions in the LGR sequences found in the sea anemones Anthopleura and Nematostella are further characterized by the occurrence of glycine- and proline-rich imperfect tandem repeats.

We generated a phylogenetic tree with all known type A LGRs and (nearly) complete predictions (Suppl. File 4). Incomplete sequences were excluded from the analysis if more than one full exon (as determined by comparison to sequences belonging to the same class of animals) was missing.

The analysis clearly shows the presence of type A LGRs throughout all animal phyla with the exception of Lophotrochozoa. In vertebrate genomes, multiple type A LGR sequences can be detected. In contrast, the analyzed invertebrates each seem to possess only one type A LGR, with the rare exception of Tribolium castaneum for which two were predicted from the genome.

Saccoglossus kowalevskii, a species of acorn worm belonging to the hemichordate phylum, has a type A LGR in its genome. This receptor’s closest relative is the LGR1 of echinoderms, generally considered to be the sister group of the Hemichordata. The urochordate Ciona has only one type A LGR in its genome. Our general phylogenetic analysis (Suppl. File 2) places this receptor in a branch next to the cephalochordate sequences. In these animals, an incomplete type A LGR was found in the lancelet Branchiostoma belcheri tsingtauense (ACM66673) and a hypothetical LGR-like protein in Branchiostoma floridae (XM_002610196).

All arthropod sequences form a separate clade, with all closely related sequences clustering together. At present, the water flea D. arenata is the only crustacean in which a type A LGR was found. Note that out of the sequences from twelve sequenced Drosophila species [13] only the sequence for D. melanogaster was added to our tree. Finally, the three nematode sequences (Caenorhabditis elegans, Caenorhabditis briggsae and Brugia malayi) and three cnidian sequences (H. magnipapillata, N. vectensis and Anthopleura elegantissima) form separate clades.

### 3.3. Type B LGRs

Type B LGRs are characterized by the presence of 16–18 LRRs, roughly twice the number found in the other types of LGRs. They have medium length hinge regions compared to the type A and type C LGRs. At the beginning of the hinge, the sequence IxxLxLYAHxCCxF is well conserved, while at the end of the hinge region the sequence IxCxPxPGFxPCEDLGSSx валют can be found in most species. Unlike type A LGRs, LGR type B receptors do not have two extra cysteines in the hinge region. Therefore, only two disulfide bridges can potentially be formed (Suppl. File 5).

As can be concluded from Supplementary Fig. 6, B-type LGRs are found in virtually all animal phyla. It is noteworthy that both predicted Strongylocentrotus and Ciona type B LGRs have shorter ectodomains and contain fewer LRRs than other type B representatives. Type B LGR sequences were also identified in the crustacean and tick species, D. arenata and ticks capulatis. These cluster together with the other arthropod type B LGRs. In fish, only complete sequences of LGR4 and LGR5 have been found. Regarding LGR5, a partial sequence was found only of Takifugu rubripes, containing the hinge region and part of the transmembrane domain (Ensembl database: ENSTRUP0000003360).

### 3.4. Type C LGRs

The total number of LRRs in type C LGRs is similar to this of the type A LGRs. One unique feature of this type is the presence of an LDLa motif, N-terminally of the LRRs. The arthropod and the chordate LGRs of type C and the sea urchin LGR7, only have a single such motif. Most interestingly however, the echinoderm (sea urchin LGR4, LGR5 and LGR8) and molluskan (Lymnaea LGR) LGRs of type C, on the other hand, harbor ten, six, five and twelve LDLa modules, respectively. A similar LGR, containing ten LDLa motifs, is present in the human body lice (Pediculus humanis corporis). From now on, we will refer to this type as type C2 LGRs. The more general type containing only one LDLa will be referred to as type C1.

The hinge region of the C-type LGRs is very short, so the other two types and the consensus sequence LxxLxHxHxPDR-FxxYcxxAPHV at its beginning and CxxPxxDGIxxEDLxxNxxVL at its end. The type C1 LGR hinges contain only two cysteines, whereas these of type C2 LGRs contain four cysteines (Suppl. File 7). In addition, some conserved amino acids differ between the two subtypes. In general, LGRs of type A, B, C1 and C2 can be distinguished by their hinge region (Suppl. File 1D).

Whereas (partial) sequences of type C LGRs were found in Aedes aegyptii (XM_001650299), Nasonia vitripennis (XM_001600282) and Acyrthosiphon pisum (XM_001942974), no type C LGR was found in
the genome of, for instance, *Bambyx mori*. In *S. kowalevskii*, one type C1 LGR has been detected so far.

An exception to the “normal” types of LGRs, are those of *Trichoplax*, the sole genus of the Placozoa, basal multicellular organisms consistent of three cell layers. In their genome, we have predicted eight LGR-like sequences, which are placed near the type C LGRs by our alignments. They do have a short hinge region, but seem to lack an LDA motif.

4. Discussion

As mentioned in Section 3, the Cnidarian genomes each hold one sequence of the type A and one of the type B LGR groups, which places the receptor’s origin above 800–1000 million years ago [70]. Based on this observation, one may hypothesize that the third type of LGRs originated later in evolution, out of another type of LGR (by gene duplication) to which an LDA encoding sequence was merged. This view, however, might be an oversimplification of the evolutionary process, since the hinge region is also different between all three types. Since the molluskan sequence analyses clearly show the presence of type CLGRs, this type had already originated before the divergence of protostomes and deuterostomes. Another factor which impedes a clear notion of the type C LGR origin are the sequences from *Trichoplax adhaerens* resembling this receptor type at the level of the hinge region. Generally, this placozoan organism is thought to be at the root of the animal kingdom, but its exact phylogenetic relationship is still a matter of debate [14].

4.1. Type A LGRs

Concerning the currently observed lack of a lophotrochozoan type A LGR, it is probably too early to draw final conclusions, given the disproportionately lower number of genome data that are currently available from this group of metazoans, although it is also quite plausible that one or more LGR types may have been lost in some branches of the evolutionary tree. This latter assumption can indeed be made in the case of the nematode *C. elegans*. This small worm has one of the first sequenced [6] and most studied genomes known, but only a single LGR has been found [36], which means that this nematode likely lost all other LGRs during evolution, as previously suggested [26]. The phylogenetic analysis considers this nematode LGR as most closely related to the type A receptors. However, detailed analysis of this receptor showed that although the hinge region clearly displays type A features, the membrane-spanning domain seems to be closer related to the type B receptors. These ambiguous properties may also account for the peculiar placement of the nematode receptors within the phylogenetic tree (Suppl. File 2).

Hitherto, in all genomes except for the vertebrate ones and this of *T. castaneum*, only one type A LGR can be identified. In vertebrates, three type A LGRs exist, namely the glycoprotein hormone receptors. These extensively studied receptors likely originate from two whole genome duplications within the vertebrate subphyllum [53], Smits and colleagues suggested that TSHR and FSHR evolved from a more recent common ancestor compared to LH/CGR, based on a mutagenesis study of TSHR and FSHR. A single substitution of the fifth amino acid of the first LRR by a leucine (L) or alanine (A) in TSHR, yielded slight sensitivity to FSH, while a single point mutation (L55Y, with Y being tyrosine) of the same residue in FSHR completely abolished sensitivity and affinity to FSH [61]. Earlier, the same conclusion was drawn from a study of several substitution mutants of the “determinant loop”, the sequence which is thought to confer binding specificity, of the ligands [7,21]. Based on the sequence similarity of the three glycoprotein hormone receptors with respect to each other and to the (single) invertebrate LGR A types, it is difficult to account for this hypothesis. While the closest relative in non-chordates (*Strongylocentrotus*) most closely resembles TSHR, other non-chordate type A LGRs seem more similar to FSHR or LH/CGR. Nevertheless, in general, they resemble all three glycoprotein hormone receptors, one slightly more than the other.

While in the tetrapod branch the known taxonomy is respected (primates cluster with primates, etc.), the bony fish show a more remarkable branching. Whereas the sequences of closely related fish such as the Nile tilapia (*Oreochromis niloticus*) and the European seabass (*Dicentrarchus labrax*), both belonging to the Perciformes, cluster together in the FSHR-branch and the TSHR-branch, they do not group together in the LHCGR-branch. The Medaka (*Oryzias latipes*) on the other hand, is in the same clade as the Nile tilapia in both the FSHR-branch as the LHCGR-branch. However, it forms a different branch altogether for the TSHR sequence. This is in part explained by recent work of Chauvigne and colleagues, who investigated the synteny of these genes in vertebrates [9]. After the bony fish specific whole genome duplication (3R) [19], all fish would have retained the same copy of the *fshr* gene, with conserved synteny to the tetrapod lineage and would have retained either one or the other copy of the *tshr* paralogues. Due to the absence of duplicates in the genome and their synteny analyses, the authors suggested that all fish retained the same copy of *lhcgr*. Therefore, it is hypothesized that the occurrence of two different *lhcgr* genes in bony fish is the result of a gene conversion [9].

Low bootstrap values within clades in all generated trees are usually the result of extremely high sequence similarity of the LGRs between species within this clade, meaning that a sequence of a particular species can branch with the sequences of many of its close neighbors. This is especially true within the vertebrate clades, where differences can be as small as a few amino acid changes. This also implies that results can get slightly distorted when an exon or some amino acids are missing in a predicted sequence.

4.2. Type B LGRs

The best characterized type B receptor is certainly the fruit fly *LGR2* or bursicon receptor [43,47]. This receptor is responsible for tanning, cuticle hardening and wing expansion after adult emergence. It is of interest to note that the type B LGRs of echinoderms (*Asterina* and *Strongylocentrotus*), mollusks (*Crassostrea*) and cnidarians (*Nematostella* and *Hydra*) are more similar to the arthropod type B LGRs than to chordate B-type LGRs (Suppl. File 6). Except for *H. magnipapillata*, which does not have a medusa stage (unlike some other members of the *Hydra* genus) these species all undergo metamorphosis from a larval stage to the adult stage. Whether B LGRs play a role during this process, analogous to the situation in insects, still remains an exciting hypothesis that requires more investigation. *Hydra* and *Nematostella*, belong to the Hydrozoa and Anthozoa respectively, two different classes of the Cnidaria that diverged from each other as much as protostomes diverged from deuterostomes [58]. This explains the relatively large distance between the sequences of these two species (Suppl. File 6).

Interestingly, the B-type receptors predicted for *Strongylocentrotus* and *Ciona* display more restricted ectodomains featuring a lower number of LRRs in comparison to other type B LGRs. The cause of this may lie in the preliminary nature of these predictions. Searches in the genome data indeed yielded additional LRRs, but assembly of the exons proved difficult, since the standard intron–exon junctions of “gt – ag” were not applicable to these putative exons. Further CDNA or EST data will be needed to overcome this obstacle.

As was discussed in the case of the type A LGRs, two genome duplication events after the branching off from the Urochordata probably tripled the sole type B LGR in invertebrates to three in
vertebrates. The similarity between the three type B LGRs is comparable to the similarity between the type A LGRs. It is of interest to note that in fish, only for *T. rubripes* an LGR5 orthologue could be identified. Perhaps, in future analyses of the genome or proteome of other fish this receptor too can be found.

Regarding the function of three B-type receptors in vertebrates, a number of studies have been published with a general tenor of their involvement in developmental processes. Research on LGR4 knockout mice revealed critical roles in embryonic growth [45], development of the male reproductive tract [46], renal development [31], keratinocyte development [32], hair formation [50], eye development [30,71] and bone formation [44]. Mice without LGR5 exhibit a full neonatal lethality, associated with ankylo glossia, the fusion of the tongue to the floor of the mouth cavity [51]. Furthermore, LGR5 and 6 have been demonstrated to be markers for epithelial stem cells [3,4]. LGR4 is localized in the small intestine where it is required for the final differentiation of Paneth cells [52].

4.3. Type C LGRs

According to our phylogenetic tree the two type C LGRs that exist in chordates (LGR7 and LGR8) resemble each other more than any other type C LGR (Suppl. File 8), with nearly 64% sequence similarity between human LGR7 and LGR8, also known as relaxin receptor 1 and 2 (RXFP1 and RXFP2). In contrast, the two fruit fly type C LGRs only display about 40% similarity, which is about the same as their similarity to the human type C LGRs. Although the three type C LGRs predicted in *B. floridana*, a very primitive chordate, do not fully match the two vertebrate types, they do share a mutual similarity of almost 70%, despite the fact that small parts of both genes have not been sequenced yet. This would be consistent with a recent gene duplication. Unfortunately, thus far no type C LGR has been identified in the genome of the sea squirt, a member of the Urochordata. Since the three LGR types have already been identified in arthropods, echinoderms and vertebrates, type C would have been expected to be present in the common ancestor of all chordates. While it is too early to draw definitive conclusions, this could mean that this type of receptors may have been lost in *Ciona*, as probably is the case for *C. elegans*.

The most closely related non-chordate in our tree, the sea urchin, a deuterostome, possesses both type C LGRs. At least three of these type C2 receptors, are present in the sea urchin genome (Suppl. File 8). This hinge region holds a motif that resembles the chordate motif and is also similar to the type A and B LGRs, albeit a lot shorter. In the body lice too, a type C2 is also present. These data might be interpreted as suggesting the possible existence of two kinds of type C LGRs (giving rise to type C1 and C2) before the divergence into proto- and deuterostomes. On the other hand, the current knowledge does not allow ruling out that this new receptor subtype has evolved independently in the respective phyla. More genomic data from other species are required to shift the discussion toward either option.

In addition, separate gene duplications have probably taken place in the chordate and arthropod lineage, since the two paralogues are more similar to each other than to their orthologues. Contrary to the type A and B LGRs, however, only two type C1 exist in the genome of vertebrates. Possibly, the paralogues of this gene were rapidly lost after the genome duplication events, retaining only two copies.

The eight LGR sequences identified from *T. adhaerens* are all very similar and, like C-type LGRs, they feature a short hinge region. However, these receptors lack an LDLa module, a C-type hallmark. Searches upstream of the predicted genes also revealed no such motifs. Therefore, the LGRs found in *Trichoplax* might be derived from the predecessors of the type C receptors, before their fusion with LDLa modules. Considering the organism’s ancient lineage, one can hypothesize that all three LGR types are derived from the genes encoding LGRs as they probably occurred in the common ancestor between Placozoa and Eumetazoa. However, extensive genomic information of *Porifera* and *Ctenophora* would be needed to investigate this hypothesis further. Alternatively, *Trichoplax* may have secondarily lost the LDLa motif [48]. It is interesting to note that previous in silico studies in *Trichoplax* also showed a very deviant gene organization for other genes, namely the pacifastins [5]. Although many annotated gene families seem to be present in *Trichoplax* [62], this organism just might prove to be an outcast with respect to most other species.

4.4. Co-evolution of LGRs and their ligands

It is clear that this subgroup of rhodopsin-related G protein-coupled receptors displays a very strong conservation throughout the animal kingdom, exemplified by the presence of at least two types of LGRs in chordate species and a third one in Placozoa. Such extraordinary preservation is indicative for its importance. A couple of years ago, Sudo et al. found a ligand for *Drosophila* LGR1 that is related to the glycoprotein hormones [63]. While this is to date, with respect to LGRs, the only clear case upon which a discussion on ligand-receptor co-evolution crossing the bridge between proto- and deuterostomes can be built, it does offer opportunities for further speculation. Apart from all vertebrate species, glycoprotein hormone related sequences can be found in the lancelet [17], the American dog tick [16], the owl limpet *Lottia gigantea* [69], the malaria mosquito, the red flour beetle, the nematode *C. briggsae*, the sea slug *Aplysia californica* and the sea urchin [68]. In addition, glycoprotein hormone sequences can be found in numerous other animals such as the desert locust *Schistocerca gregaria* [2], the hemichordate *S. kowalevskii* (LOC100370267, LOC100313754), the urochordate *Ciona intestinalis* (LOC100178300) and *Ciona savignyi* (CAR94706), the deer tick (CAR94694, CAR95342), the parasitic nematode *Trichinella spiralis* (CAR984700, CAR95342) *C. elegans* (ABK78775, CAR95352) and *B. malayi* (CAR94701, CAR95351). The mammalian glycoprotein hormone subunits and the invertebrate glycoprotein hormone subunits are probably derived from common ancestral sequences that already occurred before the divergences between ecdysozoan, lophotrochozan and deuterostome lineages [55]. While type A LGRs have been found in the genome of chordate species, no glycoprotein hormone related sequences could so far be detected in these animals using pattern and BLAST searches.

The evolution of relaxin family genes in the vertebrate lineage has already been extensively studied [20,26,28,56,74]. Syntetic mapping of the relaxin family genes, showed that all mammalian relaxin genes likely evolved from three separate ancestral genes in the common ancestor of tetrapods and teleosts: (1) AncRFLA, giving rise to INS5, (2) AncRFLB, the ancestral gene of INSL4, INSL6, RN1 and RN2, and (3) AncRFLC, which later diverged into RN3 and INS3 [68]. This last duplication took place even before the divergence of tetrapods and bony fish, indicating that the direct common ancestor already had four relaxin family genes [20]. Since RN3 was the only gene which displayed strong evolutionary constraint, comparable to insulin, it has been suggested that RN3 is the ancestral relaxin gene [20,74]. Apart from LR7, LR3 can also interact with GPRC135 and GPRC142, two non-LGR relaxin receptors [41,42]. It was proposed that LR3 is the endogenous ligand of GPRC135, based on their similar tissue distribution. It was also suggested that LR3 and GPRC135 co-evolved, since they both appear to have several gene copies in fish [10,73], and that the activation of LGR7 and LGR8 by relaxins was acquired later in mammalian evolution [73]. However, it must be noted that of the six copies found in puffer fish, only two are actual gene copies of RN3 (which arose after 3R). The other four proved to be orthologues of INS5, RN2 and INS3 [20]. Also, Park et al. showed that both relaxin 3 gene
(rln3a and rln3b, also called RFLC2 an RFLC1, respectively) products of zebrafish were able to activate both hLGR7 and hLGR8, suggesting that relaxins were already capable of activating type C LGRs prior to the appearance of mammals [56].

Relaxins belong to the insulin superfamily of peptides, which includes insulins, relaxins, insulin-like peptides (ILPs) and insulin-like growth factors (IGFs). No relaxins have unequivocally been detected in protostomian invertebrates, so far. However, it has already been suggested that ILPs might be the ligands of the invertebrate type C LGRs [75]. Many insulin-like peptides have already been detected in numerous insects belonging to different orders, such as the honey bee, locusts, mosquitoes, moths, beetles and flies [1,12,24,35,40,59,66,72,75]. ILPs can also be found in the jewel wasp (XM_001601680) and the pea aphid (XM_001950003, XM_001949218, XM_001949403). Besides insects, ILPs are found in other invertebrate phyla. For instance, three ILPs are present in Ciona [54], forty in C. elegans [29], four in the owl limpet [69] and one in the starfish Asterina pectinifera [49]. In the sea urchin, another echinoderm species, one ILP can be identified in the genome (XM_00175486). In the mollusks A. californica and L. stagnalis, at least three (AF364181, Q9NDE7, DQ479393) and four (CAA30043, AA08285, XAA09966, AA84831) ILPs can be detected, respectively. When performing a pattern search in the genome of Trichoplax one ILP can be found (XP.0002111669), which resembles Drosophila ILP6. Since no orthologues of GPCR135 or GPCR142 are present in invertebrates and since several ILPs are present in invertebrate species, it is possible that one or more of the ILPs might bind invertebrate type C LGRs. Although they lack the characteristic relaxin motif, RXXXRXXI/V, some ILPs contain part of this motif. In addition, INSIL3 (which also lacks the relaxin motif) is the endogenous ligand for LGR8 [33].

In every phylum containing a type A or a type C LGR, except for the Cnidaria, sequences related to their candidate ligands (glycophytoprotein hormones or ILPs, respectively) can be found. While this observation offers by no means full evidence for an existing co-evolution of the respective receptor/ligand pairs, it may constitute a good starting point for future research, which will likely reveal more ligand/LGR couples in a wide range of animal species.

Finally, assessing a path of co-evolution proves very difficult for the type B LGRs. In Drosophila, LGR2 was identified as the receptor for bursicon [47], a heterodimer of glycophytoprotein hormone-like cystine knot subunits which is well conserved in insects, mollusks and echinoderms, but not in chordates. In follow-up of this discovery, it was postulated that the ligand for this type of chordate receptor’s could have been replaced by another class of ligands in the course of this evolutionary process [67]. Indeed, recent studies have provided evidence for the role of R-spondins as high affinity ligands for the LGR4 and 5 receptors in mouse and human [8]. Most remarkably, while the activation of these receptors increases Wnt-dependent LRPI phosphorylation, the signal transduction does not seem to involve heterotrimeric G proteins or beta-arrestin. These results suggest there has been a change of ligand, as well as a change of signaling pathway, in the vertebrate LGR type B receptors when compared to the other receptors in this class. An interesting example of such a ligand change has recently been described for Drosophila SPR (sex peptide receptor). This GPCR is now thought to be activated ancestrally by the myoinhibiting peptides, but during the course of evolution, in Drosophila, has adopted an additional ligand, sex peptide [34,57].

5. Conclusions

This study describes and discusses the evolutionary aspects of all known and predicted LGRs. It is clear that LGRs are widely distributed and well preserved throughout the animal kingdom, which is probably a testimony of their vital importance. LGRs have now been identified (and predicted from genome data) in Placoza, Cnidaria, Nematoda, Arthropoda, Mollusca, Hemichordata, Echinodermata and Chordata. Based on their number of LRRs, their type-specific hinge region and the presence or absence of an LDLa motif, three major types of LGRs can be distinguished. Based on the number of LDLa motifs, type C LGRs can be further divided into two subtypes: a C1 type containing only one LDLa, which is present in nearly all phyla, and a C2 type containing multiple LDLa motifs and a hinge region distinct from those of the other LGR classes. So far, C2-type LGRs have only been found in echinoderms and mollusks, as well as in one insect species (P. h. corporis). In addition, eight putative LGRs were found in Trichoplax, which display similarity with type C LGRs, but lack the LDLa motif. Whether these receptors represent an ancient form of the type C LGRs, which occurred before the fusion of the LDLa motif to the rest of the receptor, or whether Trichoplax has lost this LDLa motif, remains unclear. While the evolutionary relationships of the type C LGRs seem very complex, the type A and B LGRs likely underwent two duplication events (and the loss of one of the resulting gene copies) giving rise to three members each in many species of the vertebrate lineage.

Authors’ contributions

M.B. Van Hiel has performed the sequence predictions and phylogenetic analyses and drafted the manuscript. H.P. Vandersmissen performed phylogenetic analyses and co-drafted the manuscript. T. Van Loy checked the results of all predictions and analyses and contributed to the drafting of the manuscript. J. Vanden Broeck is the senior academic author who contributed to the design of the study, the discussion of the results and the correction of the manuscript. All authors have read and approved the manuscript.

Acknowledgements

This research was financially supported by the “Belgian program on Interuniversity Poles of Attraction (IUAP/PAL P6/14)”, the Research Foundation of Flanders (FWO-Flanders) and the K.U. Leuven Research Foundation (GOA/11/002). M.B. Van Hiel obtained a Ph.D. Fellowship from the “Instituut voor de aanneming van Innovatie door Wetenschap en Technologie in Vlaanderen” (IWT) and H.P. Vandersmissen was supported by a FLOF grant (K.U.Leuven).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.peptides.2011.11.004.

References

G Model

PEP-65846: No. of Pages 8

M.B. Van Hiel et al. / Peptides (2011) xxx–xxx


