MOLECULAR EVOLUTION OF GPCRS

GLP1/GLP1 receptors

Jong-Ik Hwang, Seongsik Yun, Mi Jin Moon, Cho Rong Park and Jae Young Seong

Graduate School of Medicine, Korea University, Seoul 136-705, Republic of Korea

Correspondence should be addressed to J Y Seong
Email jyseong@korea.ac.kr

Abstract

Glucagon-like peptide 1 (GLP1) is an intestinal incretin that regulates glucose homeostasis through stimulation of insulin secretion from pancreatic β-cells and inhibits appetite by acting on the brain. Thus, it is a promising therapeutic agent for the treatment of type 2 diabetes mellitus and obesity. Studies using synteny and reconstructed ancestral chromosomes suggest that families for GLP1 and its receptor (GLP1R) have emerged through two rounds (2R) of whole genome duplication and local gene duplications before and after 2R. Exon duplications have also contributed to the expansion of the peptide family members. Specific changes in the amino acid sequence following exon/gene/genome duplications have established distinct yet related peptide and receptor families. These specific changes also confer selective interactions between GLP1 and GLP1R. In this review, we present a possible macro (genome level)- and micro (gene/exon level)-evolution mechanisms of GLP1 and GLP1R, which allows them to acquire selective interactions between this ligand–receptor pair. This information may provide critical insight for the development of potent therapeutic agents targeting GLP1R.

Key Words
- evolution
- exon
- GLP1
- GLP1R
- G protein-coupled receptor
- genome
- gene
- duplication

Introduction

Glucagon-like peptide 1 (GLP1) is a gastrointestinal peptide hormone that stimulates insulin secretion from pancreatic β-cells and promotes insulin synthesis (Schmidt et al. 1985, Fehmann & Habener 1992, Moon et al. 2011). GLP1, encoded by the glucagon (GCG) gene, is produced by tissue-specific posttranslational proteolytic processing of a large precursor containing three independent peptides GCG, GLP1, and GLP2 (Mojsov et al. 1986, Kieffer & Habener 1999). Intestinal L-cells produce mature GLP1 and GLP2 but not GCG, while pancreatic α-cells produce mature GCG but not GLP1 and GLP2. Biologically active GLP1s consist of 30 or 31 amino acids, both of which have similar bioactivities and overall metabolism, although most GLP1 secreted from the gut is C-terminally amidated and 30 amino acids in length (Mojsov et al. 1990). GLP1, secreted from L-cells in response to food intake, regulates blood glucose level by inducing insulin secretion in a glucose concentration-dependent manner, while it suppresses GCG release from pancreatic α-cells (Fehmann et al. 1995, Hansotia & Drucker 2005). The glucose-dependent action of GLP1 has potential for the treatment of diabetes mellitus because it does not cause hypoglycemia under normal plasma glucose concentration conditions (Baggio & Drucker 2002, Nauck et al. 2002). In addition to its incretin activity, GLP1 stimulates proliferation and differentiation of pancreatic β-cells (Xu et al. 1999, Drucker 2003, Li et al. 2003). GLP1 is also produced in neurons of the brainstem and is transported through axonal networks to diverse brain regions such as the cerebral cortex, hypothalamus, and thalamus.
(Vrang et al. 2007, Hisadome et al. 2010). In the brain, GLP1 is known to promote satiety by regulating appetite and food intake, resulting in body weight loss (Zander et al. 2002). GLP1 is also involved in neurite outgrowth and spatial learning ability in the central nervous system (During et al. 2003, Perry et al. 2007).

The physiological activities of GLP1 occur by interaction with the GLP1 receptor (GLP1R), which is expressed in pancreatic islets, brain, heart, kidney, and the gastrointestinal tract (Graziano et al. 1995, Wei & Mojsov 1995, Alvarez et al. 2005). Ligand-stimulated receptors increase intracellular cAMP by coupling to adenylate cyclase through the Gs protein (Drucker et al. 1987, Fehmann et al. 1995) and by elevating cytosolic calcium through activation of the Gq-mediated PLCβ pathway (Wheeler et al. 1993). However, one study reports that calcium responses may be secondary effects related to increased intracellular cAMP (Kang et al. 2001).

GLP1R is a member of the secretin receptor (SCTR)-like (class B1) family of G protein-coupled receptors (GPCRs) that comprise 15 human receptors for 20 corresponding peptide ligands (Harmar 2001, Hwang et al. 2013). As the SCTR-like receptors and their peptide ligands share similar structural features, they are likely to have emerged through gene/genome duplications from common ancestors. Duplicated receptors and peptides have undergone sequence changes during evolution, leading to diversification of their physiological functions (Acharjee et al. 2004, Kim et al. 2012). Specific diversification of peptides, i.e. conservation within orthologs but variation among paralogs, could confer selective interaction of a peptide with the cognate receptor, allowing discrimination by paralogous receptors (Wang et al. 2004, Li et al. 2005, Cho et al. 2007, Lee et al. 2009). This review describes the influence of genome/gene/exon duplications on the emergence of GLP1 and GLP1R, and the specific changes allowing selective interaction between this ligand-receptor pair.

Evolution of SCTR-like GPCRs and their peptide ligands

Since the first identification of SCTR that exhibits amino acid sequence distinct from that of other known GPCRs (Ishihara et al. 1991), the SCTR-like receptors including GLP1R have been identified in a variety of invertebrate and vertebrate species (Harmar 2001). SCTR-like GPCRs can be categorized into the following: i) corticotropin-releasing hormone receptor (CRHR) subfamily, ii) calcitonin receptor (CALCR) subfamily, iii) parathyroid hormone receptor (PTHR) subfamily, iv) GCG receptor (GCGR) subfamily including GCGR, GLP1R, GLP2R, and glucose-dependent insulinotrophic polypeptide receptor (GIPR), and v) growth hormone-releasing hormone receptor (GHRHR) subfamily including GHRHR, SCTR, vasoactive intestinal peptide receptor 1 (VIPR1), VIPR2, and pituitary adenyate cyclase-activating polypeptide receptor (ADCYAP1R1) (Harman 2001, Fredriksson et al. 2003, Oh et al. 2006, Hwang et al. 2013).

Although GLP1 and GLP1R are found only in vertebrates (Sherwood et al. 2000, Irwin & Prentice 2011), the genome history of SCTR-like GPCRs and their peptide ligand genes date back to early bilaterian ancestors (Harmar 2001, Cardoso et al. 2006). Recent advances in bioinformatic tools and public genome databases for a wide range of animal kingdoms have allowed identification of additional members and exploration of evolutionary history of the SCTR-like family (Cardoso et al. 2006, Nordstrom et al. 2008, Irwin & Prentice 2011, Hwang et al. 2013, Park et al. 2013). CRHR and CALCR subfamilies are likely the earliest members of this family as they are present in most deuterostomes as well as in the protostomes Caenorhabditis elegans and Caenorhabditis briggsae (nematode) and Drosophila melanogaster and Anopheles gambiae (arthropod) (Harmar 2001, Cardoso et al. 2006). In addition, D. melanogaster contains genes that are phylogenetically closest to CRH and CALC out of all the peptide family members (Hewes & Taghert 2001, Hwang et al. 2013). Short-range linkage analysis between protostome and deuterostome homologous regions that contain SCTR-like receptor genes demonstrated the existence of common neighbor families for SCTR-like receptor genes (Cardoso et al. 2006). Recently, the presence of protostomian PTHR-like has been demonstrated (Mirabeau & Joly 2013). Thus, it is plausible that the ancestors for the CRHR (CRHR1), CALCR, and PTHR subfamilies have emerged during early metazoan evolution before the divergence of deuterostomes and protostomes. In the deuterostome lineage, besides the CRHR, CALCR, and PTHR gene subfamilies, the GCGR and GHRHR subfamily genes occurred during chordate and vertebrate evolution. The amphioxus (Branchiostoma floridae, a representative of the cephalochordates) genome contains the CRHR, CALCR, and PTHR subfamily genes, but lacks the GCGR and GHRHR subfamily genes (Nordstrom et al. 2008, Hwang et al. 2013). Blast search of the tunicate (Ciona intestinalis, a representative of the urochordates) genome shows the emergence of the GCGR subfamily genes together with the CRHR, CALCR, and PTHR subfamily genes (Cardoso et al. 2006).
Hwang et al. 2013). However, no GHRHR subfamily genes have been identified in tunicates. In lampreys (Petromyzon marinus, a basal vertebrate agnathan), a VIPR-like gene, a member of the GHRHR subfamily genes was identified (Ng et al. 2012, Hwang et al. 2013). Thus, lamprey has SCTR-like receptors belonging to all five subfamilies.

Unlike the receptor genes, the SCT-like peptide family genes were not retrieved from nonvertebrates (protostome and early deuterostome), likely due to the considerable variation in short peptide sequences of these species (Cardoso et al. 2010, Hwang et al. 2013). In basal vertebrate lamprey, however, several peptide genes such as adrenomedullin, GHRH, VIP, ADCYAP1, and GCG were observed (Cardoso et al. 2010, Hwang et al. 2013). Despite the presence of three CRHR-like genes, the corresponding CRH-like peptide genes were not found in lamprey. As in other vertebrates, the lamprey VIP, ADCYAP1, and GCG genes may encode two or more mature peptides (Ng et al. 2012, Hwang et al. 2013). As the mature peptides have similarities in their amino acid sequences on consecutive exons, these peptides are proposed to have emerged through exon duplication events during early vertebrate evolution (Irwin et al. 1997, Sherwood et al. 2000, Irwin & Prentice 2011).

**Gene/genome duplications contribute to the emergence of GLP1 and GLP1R**

Gene families with large numbers of paralogs are established through evolutionary processes, such as gene/ genome duplications followed by gene alterations or losses (Abi-Rached et al. 2002, Larhammar et al. 2002, Holland 2003, Santini et al. 2003, Vienne et al. 2003, Kim et al. 2011, Hwang et al. 2013). In particular, two rounds (2R) of whole genome duplication during early vertebrate evolution are considered to mainly contribute to expansion of the family members. In addition, local gene duplications on a chromosome before or after 2R further diversify the family members (Hwang et al. 2013). Rapid and vast accumulation of genome sequence information of many representative taxa allows synteny-based genome comparisons and facilitates the analyses of origins and relationships of family members (Lee et al. 2009, Kim et al. 2011, 2012). Furthermore, entire genome comparisons between evolutionarily distinct taxa have led to reconstruction of hypothetical ancestral chromosomes of early vertebrates or chordates (Nakatani et al. 2007, Putnam et al. 2008). Nakatani et al. (2007), constructed hypothetical post-2R gnathostome ancestor chromosomes (GACs A0–J1 linkage groups) that were derived from pre-2R 10–13 vertebrate ancestral chromosomes (VACs A–J). Deducing locations of the gene family members to reconstructed chromosomes also provides fast and relatively accurate tools to trace gene origins (Yegorov & Good 2012, Hwang et al. 2013).

**Emergence of GLP1R**

When all SCTR-like genes were assigned on reconstructed pre-2R VAC as proposed by Nakatani et al. (2007), most genes, except for the GCGR subfamily, were found to be on the VAC E linkage group consisting of four paralogons as represented in human chromosomes 2 (GAC E0), 7 and 3 (GAC E1), 17 (GAC E2), and 12 (GAC E3) (Fig. 1). GCGR subfamily members are on different GACs (Hwang et al. 2013), which is likely due to some ancestral linkage groups being inaccurately reconstructed because of massive chromosome rearrangements in these regions, or alternatively, the gene underwent a single gene translocation that caused it to move from its authentic chromosomal fragments before or after 2R. Synteny result suggests that consecutive local duplications of the SCTR-like ancestor on VAC E generated GCGR and GHRHR ancestral genes (Fig. 1). Neighbor gene analyses of SCTR-like-containing genome fragments of the amphioxus, tunicate, and lamprey may further support this possibility. For example, 15 amphioxus SCTR-like genes are located on ten different scaffolds. Many neighbor genes in these scaffolds well pair with those on human chromosomes having the SCTR family genes (Fig. 1). In tunicates, CRHR and CALCR are closely located on chromosome 3 with neighbors that are also found on VAC E. The PTHR-like on a small scaffold also shares a common neighbor gene on VAC E. Two GCGR-like genes reside closely on chromosome 3. Interestingly, their adjacent neighbors are shown on both VAC E and GCGR family gene-containing chromosomes (Fig. 1). In lamprey, although the sizes of scaffolds containing SCTR-like genes are small so as to have very limited numbers of neighbors or no neighbors, most neighbors in these scaffolds are mapped on VAC E (Fig. 1). Interestingly, one scaffold harboring the GCGR gene subfamily has a neighbor that localizes on VAC E, while another has a neighbor on GCGR family gene-containing chromosomes. Together, these results obtained from chordates and basal vertebrate lamprey indicate a consecutive emergence of SCTR-like GPCRs during deuterostome evolution.

The 2R and local duplications, followed by sequences change and gene loss, have established osteichthyan ancestral genes for SCTR-like GPCRs and their peptide ligand (Hwang et al. 2013). Regarding GLP1R, these
neighbor gene analysis of other vertebrate species suggests families. Although different paralogons that share common neighbor gene species shows that family gene-containing genome fragments of vertebrate subbranches (Fig. 2A). Synteny analysis for and a branch comprising GCGR reveals two independent branches for GLP1R, GIPR, and GCRPR that have emerged through 2R while GCGR, GIPR, and GCRPR have emerged through 2R (Fig. 2C). Interestingly, this receptor duplication scheme is well correlated with that of corresponding peptide.

Figure 1
Conserved synteny for genome fragments containing the SCTR-like family genes. The genomic locations of the human SCTR-like family genes along with their neighbor genes were compared with those of orthologous genes in amphioxus (cephalochordate), tunicate (urochordate), and lamprey (basal vertebrate agnathan). Orthologs and paralogs are aligned on the same column with the same color. Chromosome numbers are indicated in amphioxus (cephalochordate), tunicate (urochordate), and lamprey (basal vertebrate agnathan). Orthologs and paralogs are aligned on the same column with the same color. Chromosome numbers are indicated.

by white boxes with broken lines under which ‘X’ was labeled. Paralogous genes generated by 2R exist on human chromosomes 2, 7 (or 3), 17, and 12, which fall into reconstructed linkage groups. The GCGR subfamily genes were aligned on different chromosomes as they are on different linkage groups other than GC Es. Annotations for amphioxus, tunicate, and lamprey SCTR-like family genes are shown in Hwang et al. (2013). In the case of lamprey, chromosome duplication events for generation of paralogous genes cannot be excluded although genes are on the same row.

gene/genome duplications contributed to the occurrence of GCGR, GLP1R, GLP2R, GIPR, and GCG-related peptide receptor (GCRPR) genes. GCRPR is a new member found in many vertebrate species but not in mammals (Irwin & Prentice 2011, Wang et al. 2012, Hwang et al. 2013, Park et al. 2013). The phylogenetic tree of the GCGR subfamily reveals two independent branches for GLP2R and GLP1R, and a branch comprising GCGR, GIPR, and GCRPR subbranches (Fig. 2A). Synteny analysis for GCGR subfamily gene-containing genome fragments of vertebrate species shows that GCGR, GIPR, and GCRPR are on three different paralogons that share common neighbor gene families. Although GCRPR is not found in humans, neighbor gene analysis of other vertebrate species suggests
Phylogeny and synteny for the GCGR subfamily of human (hu), mouse (mo), chicken (ch), anole lizard (an), Xenopus (xe), zebrafish (zf), medaka (md), fugu (fu), stickleback (sb), tetraodon (to), and lamprey (lam) along with human SCTR, GHRHR, VIPR, and ACDYAP1R1. The receptor sequences were aligned by using MUSCLE and a tree was constructed with MEGA 5.05. Bootstrap number indicates 100 replicates. (B) Synteny for human genome was aligned on VAC B where GLP1R is located, while the GCG ancestor is on VAC E where most SCTR-like GPCRs are found. Because the ancestors of this family, the UNC2/3, CALC, and PTH, are on VAC D, it is postulated that all these peptide members have emerged on VAC D through local duplications. Subsequently, some members may have been translocated to VAC B or E (Hwang et al. 2013). The time of GCG ancestor emergence remains unclear. Although tunicates have two GCGR-like genes, conventional Blast search tools failed to identify any peptide ligand genes for SCTR-like receptors in chordates (Cardoso et al. 2010). This is likely due to considerable amino acid variation within a short sequence of the mature peptide (Cardoso et al. 2010, Hwang et al. 2013).

Emergence of GLP1

The majority of genes for SCT-like peptide ligands are located on VAC D. The GHRH and CRH/UCN ancestors are indicated above the gene, and gene locations (megabase) are indicated below the gene. Paralogous genes are aligned on the same column with the same color. (C) Proposed evolutionary history of the GCGR subfamily genes in vertebrates. GLP2R, GLP1R, and GCRPR have emerged by local duplication before 2R while GIPR and GCRPR arose by 2R in an osteichthyan ancestor. After divergence of teleosts and tetrapods, GLP1R disappeared and GCGR were doubled through teleost-specific 3R in teleosts. The absence of the gene in each species is indicated by white boxes with broken lines.

Figure 2
Phylogeny and synteny for the GCGR subfamily genes. (A) Maximum likelihood phylogenetic tree for the GCGR subfamily of human (hu), mouse (mo), chicken (ch), anole lizard (an), Xenopus (xe), zebrafish (zf), medaka (md), fugu (fu), stickleback (sb), tetraodon (to), and lamprey (lam) along with human SCTR, GHRHR, VIPR, and ACDYAP1R1. The receptor sequences were aligned by using MUSCLE and a tree was constructed with MEGA 5.05. Bootstrap number indicates 100 replicates. (B) Synteny for human genome fragments having the GCGR subfamily genes. Chromosome numbers are

genes. In the peptide genes, GCG-, GLP1-, and GLP2-coding exons are duplicated in a single gene GCG. GCG, GIP, and GCRP arose through 2R (Irwin & Prentice 2011, Wang et al. 2012, Park et al. 2013). After divergence of teleosts and tetrapods, in teleosts GLP1R was lost and GCGR was doubled due to the teleost-specific third round (3R). There are also species-specific losses of genes. For instance, GCRPR in zebrafish and mammals and GIPR in medaka disappeared (Fig. 2C).
Because the lamprey has two GCG-like genes, the GCG ancestor is likely to have emerged before the divergence of agnathans and gnathostomes (Hwang et al. 2013).

With regard to the GLP1 family peptides, GCG (GCG–GLP1–GLP2), GIP, and a novel GCRP have emerged through 2R (Irwin & Prentice 2011, Hwang et al. 2013, Park et al. 2013). GCRP was classified as a member of the exendin-4 family in a previous study (Irwin & Prentice 2011). However, a recent study using phylogenetic analysis and biochemical ligand–receptor activity assays has revealed that GCRP was distinct from exendin-4 as well as other GLP1-related peptides (Park et al. 2013). Synteny analysis of GCG-, GIP-, and GCGR-containing genome fragments reveals similar neighbor genes, flanking these peptide family genes (Irwin & Prentice 2011). In accordance with this observation, the GCG, GIP, and GCGR genes are located on GAC E0, E2, and E3 respectively (Hwang et al. 2013, Park et al. 2013), suggesting that these three genes are ohnologous to one another. However, the fourth gene was lost before the divergence of tetrapods and teleosts.

In the teleost lineage, the 3R of genome duplication further contributed to emergence of the second form of GCG and GCGR. However, GLP1R disappeared in this lineage. The pattern of gene loss in teleosts differs by species. For example, zebrafish have lost GCRP and GCRPR, while the medaka lacks GIP and GIPR (Hwang et al. 2013, Park et al. 2013). In the tetrapod lineage, mammals have lost GCRP and GCRPR (Fig. 2C). The absence of GLP1R in teleost fish, in spite of the presence of the GLP1 peptide, remains enigmatic. With regard to the concept of coevolution of the peptide and the receptor, the possibility remains that existing receptors recognize paralogous peptides in place of the missing orthologous partners. For example, fish GLP1 binds fish GCGRs with high affinity (Irwin & Wong 2005). Thus, it is possible that GLP1 has gained a GCG-like function in teleost fish (Nguyen et al. 1994).

**Exon duplication events for GLP1 and related peptides**

Because most vertebrate GCG genes encode GCG, GLP1, and GLP2 on consecutive exons, these three peptides are likely to have emerged through exon duplication events (Fig. 3A). However, the time of exon duplication for GLP1

---

**Figure 3**

Phylogeny and exon structures of GCG and related peptide genes. (A) Exon structures of the GCG and related genes from various vertebrates and lamprey. In this diagram, as only the open reading frame-encoding exons are counted, the first exon contains the signal peptide (SP) sequence. The exons that do not encode the mature peptide sequence are also shown as simple lines. GCG, GLP1, and GLP2 are encoded by three different exons of the human GCG gene. The Xenopus GCG gene encodes three independent GLP1 peptides as well as GCG and GLP2. The second forms of medaka and zebrafish GCG genes lack GLP2. Location of the mature peptides on the exons of other peptide genes such as GIP, GCRP, ADCYAP1, VIP, SCT, and GHRH is also shown. Nonmammalian ADCYAP1 and VIP encode PACAP-related peptide (PRP) and VIP-related peptide (VRP), respectively, while those sequences in human genes are dysfunctional. The vestiges that remained in the human genes are indicated by spaces covered with gray dashed lines. (B) Maximum likelihood phylogenetic tree for GLP1-related peptides of human (hu), mouse (mo), chicken (ch), anole lizard (an), Xenopus (xe), zebrafish (zf), medaka (md), fugu (fu), stickleback (sb), tetraodon (to), and lamprey (lam) along with human SCT, GHRH, VIP, and PACAP are shown. The mature peptide sequences were aligned using MUSCLE, and a bootstrap consensus tree was constructed using MEGA 5.05. Bootstrap numbers represent 100 replicates.
is unclear (Sherwood et al. 2000). The basal vertebrate lamprey has two GCG-like genes (Irwin et al. 1999). One (lam1GCG) encodes GCG and two GLP2-like peptides (GLP1/2 and GLP2), and the other (lam2GCG) encodes GCG and GLP2. The phylogenetic position of lamGLP1/2 is currently unclear. A previous study using Clustal X-2.1 alignment followed by the neighbor-joining phylogenetic tree indicated that lam1GLP1/2 is more similar to GLP1 than to other related peptides (Hwang et al. 2013), while this study using multiple sequence alignment with high accuracy (MUSCLE) alignment with a maximum likelihood phylogenetic tree shows a close relationship of lam1GLP1/2 with mammalian GLP2 (Fig. 3B). Despite the confusion, this observation suggests that exon duplication events date back to early vertebrate evolution before the divergence of agnathans and gnathostomes. Owing to the short nucleotide sequences of scaffolds (GL481103 and GL477955) harboring the lamprey GCGs, it is still uncertain whether the two lamprey GCGs were generated by genome duplication or local gene duplication. However, as very close relationships between peptides encoded by lam1GCG and lam2GCG are evident from the phylogenetic tree (Fig. 3B), these two genes may have emerged recently, probably through local duplication. Interestingly, the ADCYAP1 and VIP genes in vertebrates and basal vertebrate lamprey, which are believed to have emerged next to the GCG gene, also encode two mature peptides, and one (PACAP-related peptide (PRP) or VIP-related peptide (VRP)) of these two peptides is similar to GHRH in amino acid sequence (Laburthe et al. 1996, Sherwood et al. 2000, Ng et al. 2012, Hwang et al. 2013). Although the physiological roles of PRP and VRP remain incompletely understood, their interactions with specific receptors have been described in zebrafish and chicken (Wang et al. 2010). In addition, some vertebrate GHRH genes also encode PACAP- or VIP-like peptide sequences, although these peptides are likely nonfunctional. It is important to note that the PACAP, VIP, GHRH, and SCT peptide sequences are relatively more closely related to the GLP2 peptide when compared with other related peptides (Fig. 3B). This observation raises a possibility that the ancestor for ADCYAP1, VIP, GHRH, and SCT emerged through local gene duplication of the GLP2-containing ancestral gene before 2R, and that exon duplication events in this ancestor may influence the presence of the two peptides in the ADCYAP1, VIP, and GHRH genes (Fig. 3B). The exon duplication event before 2R was followed by additional exon duplication in Xenopus of the tetrapod lineage, or exon degeneration after 3R in the teleost lineage. The GCG gene of Xenopus, a representative amphibian, contains two more exons encoding GLP1s. A copy of teleost-specific duplicated GCGs has lost the GLP2 sequence. In zebrafish, both GCGs lack the GLP2 sequence (Fig. 3A).

Specific amino acid sequence changes determine selective ligand–receptor interactions

GLP1R and its paralogous receptors exhibit high degrees of amino acid sequence identity with one another and share a relatively long N-terminal extracellular domain (ECD) beginning with an α-helix followed by four β-strands forming two antiparallel sheets that are stabilized by disulfide bonds and a salt bridge (Grace et al. 2004, Parthier et al. 2007, Runge et al. 2008, Underwood et al. 2010). Likewise, all peptide ligands for these receptors have typical three-dimensional structures to facilitate receptor binding. GLP1 and its family of peptides consist of 30–40 amino acids which share similarities in amino acid sequence and secondary structure. The structures of the peptides demonstrated by nuclear magnetic resonance and crystallography reveal that the N-terminal six or seven amino acid residues of peptides form a random coil structure where they share more than 70% amino acid sequence identity, and this is followed by an α-helical structure with moderate sequence similarity among the peptides (Thornton & Gorenstein 1994, Neidigh et al. 2001, Alana et al. 2007, Parthier et al. 2007, Underwood et al. 2010). Some residues with different lengths are extended after the α-helix to form flexible regions in the C-terminus, although this region is not necessary for receptor recognition and activation (Hinke et al. 2003, Moon et al. 2010). These peptides form a common local structure, called a helix N-capping motif, to stabilize the helical structure (Neumann et al. 2008). This motif in GLP1 is formed between hydrophobic amino acids Phe-6 and Val-10, and the polar amino acid Thr-7 through hydrophobic interaction and hydrogen bonding. These interactions are believed to facilitate receptor activation upon ligand–receptor binding (Fig. 4A). While the N-terminus and the first half of the central α-helical structures are responsible for interaction with the receptor core domain, the second half of the α-helical region binds directly to the ECD of the receptor (Parthier et al. 2007, Underwood et al. 2010, Moon et al. 2012a).

Interaction between peptide N-terminus and the receptor core domain

Although the GLP1 peptide and receptor families exhibit relatively high sequence similarities and conserved
structures (Sherwood et al. 2000), each peptide ligand selectively interacts with its own cognate receptor, indicating that this diversification is particularly important for selective interaction of the ligand–receptor pair, resulting in the specific physiological functions of each peptide-receptor pair (Moon et al. 2012b, Park et al. 2013).

In the N-terminal part of GPL1, residues His-1, Gly-4, Phe-6, Thr-7, and Asp-9 are crucial either for maintaining secondary structure of the peptide or for interaction with the receptor (Adelhorst et al. 1994, Gallwitz et al. 1994). Particularly, Gly-4, Phe-6, and Asp/Glu-9 are highly conserved for all GLP1 and related peptides (Fig. 4B), such that these residues in the related peptides are commonly important for recognition of their cognate receptors, particularly the core domain of the receptor (Perret et al. 2002, Hinke et al. 2003, Runge et al. 2003a, Yaqub et al. 2010). His-1 and Thr/Ser-7 are conserved in the GLP1 family of peptides (GLP1, GLP2, GCG, and GCRP) except for GIP that has Tyr-1 and Ile-7 in the corresponding sites (Fig. 4B), raising a possibility that His-1 and Thr-7 of GLP1 are likely important for selective interaction with GLP1R, but discriminate GIPR. Indeed, introducing His-1 and Thr-7 into GIP allows this chimeric peptide to bind and activate GLP1R with a relatively high affinity and potency (Moon et al. 2010, 2012a). In contrast, Tyr-1 and Ile-7 in GIPR are important for selective interactions with its receptor. In addition to Tyr-1 and Ile-7, alanine scanning or amino acid modification studies reveal that highly conserved residues Glu-3, Gly-4, Asp-9, and Asp-15 are found to be important for receptor activation (Hinke et al. 2003, Yaqub et al. 2010). In GCG, besides the highly common residues Gly-4, Phe-6, The-7, and Asp/Glu-9 and GCG-specific residues Ser-2, Gln-3, Tyr-10, and Lys-12 (Fig. 4B) are found to be important for receptor binding and activation (Perret et al. 2002, Runge et al. 2003a). Interestingly, introducing these individual residues into the equivalent positions of GLP1 did not significantly alter peptide affinity and potency towards GLP1R. However, the replacement of these residues in GCG with those of GLP1 greatly reduced GCR affinity.
(Runge et al. 2003a), indicating that positions 2, 3, 10, and 12 are only important for GCG. This result is consistent with the observation that wild-type (WT) or chimeric peptides having either the GCG or GLP1 N-terminus interact with receptors having the GLP1R core domain while WT or chimeric peptide containing the GLP1 N-terminus poorly binds or activates receptors containing the GCGR core domain (Runge et al. 2003b). The structure–activity relationships of GLP2 have not been extensively analyzed. However, an alanine scanning analysis revealed that residues at positions 2, 5, 6, and 17 were important for receptor activation (DaCambra et al. 2000).

A recent observation, after comparing amino acid sequences of the GLP1 peptide family including a novel GCRP, indicates that residues at positions 16–18 exhibited ortholog-specific amino acid changes, such that the Lys-16–Met/Ile-17–Lys-18, Ser/Thr-16–Arg-17–Arg-18, and Gly/Asp-16–Gln-17–Ala-18 motifs were conserved in GCRP, GCG, and GLP1 respectively (Park et al. 2013). Interestingly, placing the Gly/Asp-16–Gln-17–Ala-18 motif in GCRP increased peptide potency toward the WT GLP1R and GCRPR chimeric receptor that has the GLP1R core domain (Park et al. 2013). This observation raises the possibility that the motif at positions 16–18 may contribute to selective interactions with their cognate receptors. This motif resides at the border between residues that interact with the ECD and residues that contact the core domain. Particularly, Gly-16 of GLP1 contributes to the formation of a kink in the peptide helix (Underwood et al. 2010). Thus, this motif can be proposed to be involved in interaction between ECD and the core domain under ligand-bound conditions. A recent study has proposed that the ECD of GCGR is likely to negatively regulate GCGR activity through interaction with extracellular loop 3 (ECL3) of the core domain (Koth et al. 2012).

As ligand-bound crystal structures of the core domain of the GLP1R family are not available, the residues in the core domains responsible for ligand binding or receptor activation are poorly understood, although many approaches, such as alanine scanning, photoaffinity labeling, and molecular docking modeling have been attempted to identify these specific amino acid residues (Xiao et al. 2000, Lopez de Maturana & Donnelly 2002, Lopez de Maturana et al. 2004, Chen et al. 2009, 2010, Lin & Wang 2009). Mutation mapping studies suggest that charged amino acids (Lys-197, Asp-198, Lys-202, Asp-215, Arg-227, and Lys-288) and conserved residues (Met-204, Tyr-205, and Thr-306) are likely important for GLP1-induced receptor activation (Xiao et al. 2000, Lopez de Maturana & Donnelly 2002, Al-Sabah & Donnelly 2003, Lopez de Maturana et al. 2004, Koole et al. 2012). However, these studies do not explain how individual residues of the peptide interact with residues in the receptor. Our recent study reveals that His-1 and Thr-7 were found to interact with Asn-302 at ECL2 and Ile-196 at the upper half of transmembrane helix 2 (TMH2) in GLP1R respectively (Moon et al. 2012a). Interestingly, Asn-302 is conserved for all vertebrate GLP1R, and at corresponding positions in GCGR (Asn-300 for human), GLP2R (Asn-336 for human), and GCRPR (Asn-301 for chicken), that are activated by peptide ligands containing His-1. GIPR, however, has hydrophobic Val-293 at this position (Moon et al. 2012b). Likewise, all receptors interacting with Thr-7-containing peptides have conserved hydrophobic Ile or Val at the corresponding position of Ile-196 in GLP1R, while GIPR has Ser-189 in this position. The study using photolabile probes of GLP1 provided only partial information on spatial approximation of interacting residues of the peptide and receptor (Chen et al. 2010). Molecular docking modeling with biochemical analyses has proposed the ligand-binding pockets in the GLP1R core domain (Lin & Wang 2009, Coopman et al. 2011, Kirkpatrick et al. 2012). However, these modeling approaches failed to provide a common consensus of the ligand-binding pocket such that ligand contact residues in the binding pocket are all different from one another. Thus, stricter biochemical analyses with an appropriate strategy may be helpful to construct a more reliable ligand-bound receptor model.

Interaction between the peptide C-terminus and the receptor ECD

Interactions between the peptide C-terminus and the receptor ECD are well defined for GLP1R and GIPR as demonstrated by ligand-bound ECD crystal structures (Parthier et al. 2007, Runge et al. 2008, Underwood et al. 2010). In addition, these crystal structures were useful in understanding interactions between GCG and the GCGR ECD and between GLP2 and the GLP2R ECD by applying molecular docking models (Venneti & Hewage 2011, Koth et al. 2012, Siu et al. 2013). The crystal structure of the ligand-bound ECD of the receptors shows that only residues in the second half of the α-helix (ranging from Ala-18 to Val-29 for GLP1) interact with the ECD (Underwood et al. 2010). The α-helices of all peptide hormones are partly amphipathic allowing hydrophobic and hydrophilic interactions with ECD of their receptor. The hydrophobic face (Ala-18, Ala-19, Phe-22, Ile-23, Trp-25, and Leu-26) of GLP1 is toward the hydrophobic binding cavity of the ECD that is formed by residues...
Leu-32, Thr-35, Val-36, and Trp-39 in the N-terminal α-helix, residues Tyr-69 at the end of the β2 strand, and residues Try-88, Leu-89, Trp-91, and Pro-90 in the loop between the β3 and β4 strands (Runge et al. 2008, Underwood et al. 2010). Many of the residues in the binding pocket are highly conserved in the equivalent positions of GIPR, GLP2R, and GCGR and they are found to be important for peptide binding (Parthier et al. 2007, Venneti & Hewage 2011, Koth et al. 2012). Considering that residues Ala-19, Phe-22, Ile-23, and Leu-26 in the hydrophobic face of the peptides are also highly conserved, this result suggests that these evolutionarily conserved residues in the peptides and receptors are likely to contribute to primary binding between the α-helical domain of the peptides and the ECD of the receptors (Parthier et al. 2007, Underwood et al. 2010, Venneti & Hewage 2011, Koth et al. 2012). This primary interaction may explain cross-interactions of peptides with other receptors in the same family. For example, the GLP1 chimeric peptides having the α-helical region of GIP or GCG are able to activate GLP1R with a relatively high potency (Runge et al. 2003a,b, Moon et al. 2010). The hydrophilic face of GLP1 comprises residues Gln-17, Lys-20, Glu-21, and Lys-28. Of these residues, only Lys-20 was found to make a direct interaction with the ECD residue Glu-128 (Underwood et al. 2010). At positions 17–20, including the hydrophilic residues, GLP1 and its related peptides exhibited marked variations in amino acid sequence, raising a possibility that selective recognition of the peptide toward the cognate ECD can be achieved through interactions of these ortholog-specific residues of the α-helix with specific residues in the ECD of the cognate receptor (Parthier et al. 2007, Underwood et al. 2010).

**Conclusions**

GLP1 and GLP1R have emerged through 2R and local duplications before and after 2R. The exon duplication events also contributed to occurrence of the GLP1 peptide. Because the GLP1 peptide and receptor families arose from a common ancestor, they share similarities in their amino acid sequences and basic structures. Their conserved residues and structural similarities across the family members suggest a common mechanism underlying ligand–receptor binding followed by receptor activation. The comparison of ligand–receptor interaction mechanisms reveals that the conserved residues among GLP1 and related peptides tend to interact with conserved residues of GLP1R and its paralogous receptors (Moon et al. 2012b). Specific change after genome/gene/exon duplication may allow selective interaction of each ligand–receptor pair, and could provide for diversified functions. Our recent studies have demonstrated that interactions between residues evolved from ortholog-specific peptides and receptors enhanced recognition of a peptide for the cognate receptor, while maintaining discrimination by other receptors (Moon et al. 2010, 2012a). During the evolutionary process, GLP1, through its specific interaction with GLP1R, acquired a particular function in glucose homeostasis and appetite regulation. Although GLP1 has gained great attention due to its beneficial effects on type 2 diabetes and obesity, molecular interaction with its receptor is not fully understood. Particularly, the specific amino acid residues within the TMH and ECL of receptors that may confer ligand binding and receptor activation remain poorly investigated. Understanding the evolutionary history of GLP1/GLP1R can provide clues to explore molecular interaction properties between GLP1 and GLP1R. Elucidation of the structure of the ligand-binding pocket in the core domain of GLP1R may facilitate development of potent peptide agonists or small molecules capable of activating GLP1R, which could greatly contribute to the treatment of diabetes and obesity.

**Declaration of interest**

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the review.

**Funding**

This work was supported by grants from the Brain Research Program (2011-0019205) and Basic Research Program (2010-0022054) of the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning.

**References**


Yegorov S & Good S 2012 Using paleogenomics to study the evolution of gene families: origin and duplication history of the relaxin family hormones and their receptors. PLoS ONE 7 e32923. (doi:10.1371/journal.pone.0032923)


Received in final form 16 October 2013
Accepted 10 December 2013