



Insight into Evolution and Conservation Patterns of B1-Subfamily Members of GPCR

Chiranjib Chakraborty^{1,2} · Ashish Ranjan Sharma² · Garima Sharma³ · Manojit Bhattacharya² · Sang-Soo Lee²

Accepted: 30 January 2020 / Published online: 8 February 2020
© Springer Nature B.V. 2020

Abstract

The diverse, evolutionary architectures of proteins can be regarded as molecular fossils, tracing a historical path that marks important milestones across life. The B1-subfamily of GPCRs (G-protein-coupled receptors) are medically significant proteins that comprise 15 transmembrane receptor proteins in *Homo sapiens*. These proteins control the intracellular concentration of cyclic AMP as well as various vital processes in the body. However, little is known about the evolutionary correlation and conservational blueprint of this GPCR subfamily. We performed a comprehensive analysis to understand the evolutionary architecture among 13 members of the B1-subfamily. Multiple sequence alignment analysis exhibited six multiple sequence aligned blocks and five highly aligned blocks. Molecular phylogenetics indicated that CRHR1 and CRHR2 share a typical ancestral relationship and are siblings in 100% bootstrap replications with a total of 24 nodes observed in the cladogram. CRHR2 has the maximum number of extremely conserved amino acids followed by ADCYAP1R1. The longest continuous number sequence logos (74) were found between sequence location 349 and 423, and consequently, the maximum and minimum logo height recorded was 3.6 bits and 0.18 bits, respectively. Finally, to understand the model and pattern of evolutionary relatedness, the conservation blueprint, and the diversification among the members of a protein family, GPCR distribution from several species throughout the animal kingdom was analysed. Together, the study provides an evolutionary insight and offers a rapid method to explore the potential of depicting the evolutionary relationship, conservation blueprint, and diversification among the B1-subfamily of GPCRs using bioinformatics, algorithm analysis, and mathematical models.

Keywords Secretin receptor · B1-subfamily · In silico · Evolutionary relationship

Chiranjib Chakraborty and Ashish Ranjan Sharma have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10989-020-10043-5>) contains supplementary material, which is available to authorized users.

✉ Chiranjib Chakraborty
drchiranjib@yahoo.com

✉ Sang-Soo Lee
123sslee@gmail.com

¹ Adamas University, North, 24 Parganas, Kolkata 700126, West Bengal, India

² Institute for Skeletal Aging & Orthopedic Surgery, Chuncheon Sacred Heart Hospital, Hallym University, Chuncheon 24252, Republic of Korea

³ Neuropsychopharmacology and Toxicology Program, College of Pharmacy, Kangwon National University, Chuncheon 24341, Republic of Korea

Introduction

The name “protein family” was proposed by Dayhoff in the 1960s to describe proteins with similar structure and/or function, which have followed the evolution process from a common ancestor protein (Dayhoff 1969). The GPCR family are physiologically significant membrane proteins that are associated with different signalling pathways and represent one of the most significant target-classes for drug discovery. Two important papers were published about GPCRs in *Nature* and *Nature Review Cancer* describing the molecular signatures (Venkatakrishnan et al. 2013) and mutational landscape of GPCRs (O’hayre et al. 2013). Several papers were published about GPCRs in *Nature* and *Cell* group journals demonstrating the importance of this family. GPCRs were classified into six families which are- (i) secretin-receptor family, (ii) rhodopsin family, (iii) metabotropic glutamate receptor family, (iv) fungal pheromone A- and M-factor receptors, (v) cyclic-AMP receptors, and (vi)

fungal pheromone P- and α -factor receptors from *Dictyos-
telium* (Kolakowski Jr 1994). The secretin receptor family, a
well-known member of GPCRs, abundantly expression are
found in acinar cells as well as ductal epithelial cells, with
negligible levels within pancreatic vascular structures as
well as in islets (Siu et al. 2006; Ulrich et al. 1998b). These
family members are most abundant in the liver (Alpini et al.
1994), stomach (Bawab et al. 1988) (Gespach et al. 1981; Li
et al. 1998), brain (Freneau et al. 1983; Köves et al. 2002)
(Nozaki et al. 2002; Yung et al. 2001), reproductive system
(Chow et al. 2004), intestine (Andersson et al. 2000), heart
(Ishihara et al. 1991), along with lungs (Christophe et al.
1981). GPCRs have extreme therapeutic potential. These
receptors are presently used as drug targets of several dis-
eases (Araç and Leon 2020; Purcell and Hall 2018; Vass
et al. 2018).

The secretin family, also known as the ‘B family’ (Car-
doso et al. 2005; Hamby and Hirst 2008), consists of classi-
cal receptors (Cardoso et al. 2005). B family is grouped into
three subfamilies: B1, B2, and B3 (Kolakowski Jr 1994).
The B1 family consists of 15 protein members (Kolakowski
Jr 1994; Cardoso et al. 2014). In our study, we analysed the
evolutionary relatedness of 13 members of B1 subfamily of
the secretin receptor family, including calcitonin receptor
(CALCR), adenylate cyclase-activating polypeptide 1 (pitu-
itary) receptor (ADCYAP1R1), corticotrophin-releasing hor-
mone receptor 2 (CRHR2), gastric inhibitory polypeptide
receptor (GIPR), corticotrophin-releasing hormone recep-
tor 1 (CRHR1), glucagon-like peptide receptor 1 (GLP1R),
growth hormone-releasing hormone receptor (GHRHR),
glucagon-like peptide receptor 2 (GLP2R), parathyroid hor-
mone receptor type 1 (PTH1R), secretin receptor (SCTR),
parathyroid hormone receptor type 2 (PTH2R), two type
of vasoactive intestinal peptide receptor type 1 (VIPR1) and
type 2 (VIPR2). However, and we have are not considered
calcitonin receptor-like receptor and glucagon receptor in
this study due to inadequate data availability.

The adenylate cyclase-activating polypeptide type I
(pituitary) receptor, also entitled as PAC1, ADCYAP1R1,
PAC1R, PACAPR, or PACAPRI, is a protein which is
expressed from the *ADCYAP1R1* gene in humans. The
protein binds with another peptide entitled “pituitary
adenylate cyclase-activating peptide (PACAP)” (Ogi et al.
1993; Vaudry et al. 2000). Calcitonin receptor (CALCR)
is occupied with the keeping with the regulation of osteo-
clast-mediated bone resorption and calcium homeostasis.
However, the onset of osteoporosis and variations in bone
mineral density are associated with the polymorphisms
in this gene (Chambers and Magnus 1982; Dacquin et al.
2004; Davey et al. 2008). Corticotrophin-releasing hor-
mone receptor is made up of two types of receptors:
CRHR1 and CRHR2 and they are expressed by the *CRHR1*
and *CRHR2* genes, correspondingly. This receptor binds

to corticotropin-releasing hormone (CRH) (Hauger et al.
2003). CRHR1 attaches with neuropeptides of the CRH
family of proteins. These proteins are the central controller
of the hypothalamic–pituitary–adrenal pathway. CRHR1
is an essential protein that can activate of signal trans-
duction pathways, and it controls various processes which
include immune response, stress, obesity, and reproduc-
tion, while CRHR2 shows extreme attraction for CRH, and
can attach with CRH-associated peptides (e.g. urocortin).
It has been noted that CRH is produced in the hypothala-
mus which acts as a vital responsibility to coordinate the
autonomic, endocrine, and behavioural responses to pres-
sure and immune challenge (Aguilera et al. 1986) (Gram-
matopoulos and Chrousos 2002). Gastric inhibitory poly-
peptide receptor (GIPR) is encoded by the *GIPR* gene,
which produces a GPCR which is GIP (gastric inhibitory
polypeptide) and it hinders gastric acid secretion and
gastrin release in addition to stimulating insulin release
in the presence of elevated glucose (Stoffel et al. 1995).
Glucagon-like peptide receptor consists of type 1 receptor
(GLP1R) and type 2 receptor (GLP2R). GLP1R is secreted
from pancreatic beta cells, and helps in stimulate GLP1R
activates the adenylyl cyclase pathway; It affects increased
insulin synthesis and liberates of insulin (Drucker et al.
1987). As a result, GLP1R is a possible target for the man-
agement of diabetes (Holmes 2003). GLP1R found in the
brain, where it is associated with appetite control (Kinzig
et al. 2002). The proglucagon peptide synthesises GLP2R.
The *GCG* gene is expressed in this peptide. GLP2/GLP2R
helps in augmenting the villus height in the small intestine,
stimulates intestinal growth, reduce enterocyte apoptosis,
and associated with increased crypt cell propagation. Fur-
thermore, GLP2/GLP2R prevents intestinal hypoplasia,
and that is the consequence of total parenteral nutri-
tion. *GHRHR* gene helps in the expression of the growth
hormone-releasing hormone receptor (GHRHR). Simi-
larly, the *GHRH* gene helps in the expression of growth
hormone-releasing hormone (GHRH). The occurrence
of receptor association of GHRH causes the production
and liberate of growth hormone. Growth hormone defi-
ciency (IGHD), also known as “Dwarfism of Sindh,” is
caused by the mutations in the *GHRH* gene. This disorder
is categorised by short stature (Murray et al. 2000). Para-
thyroid hormone receptor has two sub-types: parathyroid
hormone receptor type 1 (PTH1R) and type 2 (PTH2R).
The PTH1R gene encodes type 1. PTH1R is a receptor
for two peptides, which are parathyroid hormone-like
hormone (PTH1H) and parathyroid hormone (PTH). Any
deficiency in PTH1R is known to cause different diseases
such as Chondrodysplasia Blomstrand type (BOCD), Jans-
en’s metaphyseal chondrodysplasia (JMC), and enchon-
dromatosis (Calvi and Schipani 2000). The *PTH2R* gene
encodes the type 2 receptor. It is more specific in ligand

detection. This gene has a tissue-specific delivery. PTH1R makes active only through the PTH but not by PTHLH. This protein is predominantly available in the pancreas and brain (Bhattacharya et al. 2011). Through the binding secretin, SCTR (Secretin receptor) is activated, and it is the powerful regulator of pancreatic electrolyte, pancreatic bicarbonate etc. SCTR, along with Secretin, may be associated with autism and pancreatic cancer (Dong and Miller 2002) (Felley et al. 1992). Two receptors are associated with the category of the vasoactive intestinal peptide receptor, which are vasoactive intestinal peptide receptor type 1 (VIPR1) and type 2 (VIPR2). These receptors are connected with the secretion of exocrine and endocrine, smooth muscle relaxation, ion and water and flux in lung as well as epithelial cell in intestine. Their functions are linked with integral membrane receptors. This receptor acts through the guanine nucleotide-binding. Finally, it triggers adenylate cyclase (Laburthe and Couvineau 2002).

Evolution involves the genetic variations among the existing individuals in a population. These variations arise as a result of the fundamental biological process of replication to form DNA, which is then transcribed to RNA and finally translated to create proteins with the help of ribosomes. Thus, the protein amino acid sequence is analysed by researchers to predict the structural and functional protein units through conserved patterns. These highly conserved regions are directly involved in protein–protein interaction (Branden 1999). Researchers can also gain information on protein folding, function, and transport through protein glycosylation, as these conserved regions are essential for the cell–cell interaction and antigenicity (Hamby and Hirst 2008). A protein can be evaluated for evolutionary conservation by comparing the position of functionally and structurally essential amino acids. Therefore, conservation analysis of amino acid residue positions among the family members can uncover the significance of every location for the structure as well as the function of a particular protein (Ashkenazy et al. 2010; Chakraborty and Agoramoorthy 2014). In evolutionary biology, it is essential to know the delivery of a specific family of protein in each species (Herrada et al. 2011), and several proteins have been analysed in such a context to understand their distribution (Atkinson et al. 2011; Peng et al. 2012).

The goal of this *in silico* analysis is to gain insight into the evolutionary architecture and conservation blueprint among different members of the B1-subfamily of the secretin receptor proteins. The aim of this study was (i) to analyse the multiple sequence alignment and their score determination, (ii) to investigate the phylogenetic relationship among the member proteins, (iii) to understand the conservation patterns of known highly conserved amino acids, and (iv) to depict the distribution of B1-subfamily members across different species.

Materials and Methods

Data Acquisition

There are 15 protein members in the B1-subfamily of the secretin receptor family, and in our study, 13 members have been analysed. Due to inadequate data, two members, calcitonin receptor-like receptor and glucagon receptor, have not investigated. All of the data for the genes related to protein receptors of the B1-subfamily of the secretin receptor family, including ADCYAP1R1, CALCR, CRHR1, CRHR2, GLP1R, GLP2R, GIPR, GHRHR, SCTR, PTH1R, PTH2R, VIPR1, and VIPR2, were collected from the NCBI protein database (Wheeler et al. 2007). The protein sequence of all the proteins was obtained from NCBI in FASTA format and analysed for suitability for the research study. We analysed the distribution of the secretin family protein members (Family: 7tm_2 (PF00002)) in different species and different kingdoms, using the Pfam server, which contains a database of protein families and domains.

Multiple Sequence Alignment (MSA) and Score Determination

Sequences were analysed by performing multiple sequence alignment using the CLUSTAL W (version 2.1) server (Chenna et al. 2003), and sequence similarities were observed. We performed multiple sequence alignments using CLUSTAL W (version 2.1) with 14 sequences belonging to nine members of the B1-subfamily. The alignment score is calculated and used for further analysis. The MSA was compacted into one by one performing each profile alignment using MUSCLE software (Edgar 2004). MUSCLE employs a novel function, and it can be elucidated as through log-expectation score function as follows:

$$LE^{xy} = (1 - f_G^x)(1 - f_G^y) \log \sum_i \sum_j f_i^x f_j^y p_{ij} / p_i p_j \quad (1)$$

which is a customized edition of the log-average function and is as follows:

$$LA^{xy} = \log \sum_i \sum_j f_i^x f_j^y p_{ij} / p_i p_j \quad (2)$$

where i and j are amino acid category; p_i is the probability of the condition of i ; p_{ij} is the joint probability or standard probability of i and j notation which are aligned to each other; f_i^x is the frequency (observed frequency) of i in column x of the initial profile; and f_G^x the frequency (observed frequency) of spaces/gaps in that column at position x in the family and similarly for position y in the second profile. The

approximate probability is α^x , where observing amino acid i location and x can be calculated from fx .

Again, we applied the Gblocks server (version 0.91b), to examine the aligned blocks, which defines a position of conserved blocks from numerous alignments (Aguilera et al. 1986; Castresana 2000). All the above servers use maximum likelihood algorithms (Pupko et al. 2002), or they use the empirical Bayesian algorithms (Mayrose et al. 2004) to study the MSA.

Phylogenetic Tree Building as well as Bootstrap Estimation Study

A phylogenetic tree is used to show the evolutionary relatedness based on evolutionary divergence. We used the results generated through sequence alignment to fabricate the phylogenetic tree (Phylogram) using computational bioinformatics server (Phylogeny.fr) (Dereeper et al. 2008), which uses essential algorithms from MUSCLE, PhyML, TreeDyn and Gblocks. All of these servers utilize the ‘bottom-up clustering algorithm’ ‘neighbour-joining algorithm’. This algorithm takes the input as well as distance matrix to indicate the divergence/distance among every couple of taxa, and this matrix has a dimension of $N \times N$. In this case, N is the quantity of points in number or number of nodes. Lastly, we performed bootstrap analysis (Hillis and Bull 1993; Holmes 2003) using the developed phylogenetic tree to make inferences and evaluate robustness.

Analysis of Cladogram Using Computational Biology and Unrooted Tree Formation

From Phylogram, we have constructed a cladogram, binary tree, and un-rooted tree. The binary tree represented the cladogram as the number of nodes in each different level. Again, two types of the unrooted phylogenetic tree were developed, including unrooted-alpha and circular alpha. MAFFT (version 6.0) was applied to construct the unrooted-alpha phylogenetic tree (Katoh and Toh 2008). Interactive Tree Of Life (iTOL) was utilised for circular alpha phylogenetic tree creation (Letunic and Bork 2006).

Conservation Patterns Examination and Computation of Extremely Conserved Amino Acids

The ConSurf server was used to predict the conservation models (Ashkenazy et al. 2010; Glaser et al. 2003). The evolutionary concaveness of amino acids was calculated in each position of amino acid in every ‘B1 subfamily’ proteins using a Bayesian algorithm. The score

of conservation at each amino acid is generated with the application of the same tool, and finally, the highly conserved amino acids were examined using the same tool.

Sequence Logos of Conserved Domains

The WebLogo software is used to generate a sequence logo. This is utilized for a graphical illustration of amino acid sequences and this tool is displaying the prototype as a bunch of aligned sequences (Crooks et al. 2004) (Schneider and Stephens 1990). The software is used to examine the aligned pattern and profile of bias across amino acid.

Schneider et al. (Schneider and Stephens 1990) described the sequence logo at a particular position which can be characterised as R_{seq} , which is as follows:

$$R_{seq} = S_{max} - S_{obs} = \log_2 N - \left(\sum_{n=1}^N p_n \log_2 p_n \right) \quad (3)$$

where the difference between the utmost probable entropy and the entropy of the observed sample distribution is represented as R_{seq} ; The studied frequency of symbol n at a particular sequence position is denoted as p_n and the number of specific symbols for the agreed sequence type is represented as N .

Distribution of B1-Subfamily Proteins Among Different Species

To comprehend the distribution of proteins in ‘B1 subfamily’ throughout diverse species, we have used the Pfam software (Punta et al. 2011), which is a massive collection of families of protein and their domains. This server has an inbuilt characteristic that can search the database of protein families and domains. It produces the distribution of the protein family members among the different species, as well as the various super kingdoms. Finally, it depicts the result as (i) ‘‘sunburst’’ tree and (ii) ‘‘species tree,’’ as well as (iii) ‘‘the phylogenetic tree for family seed alignment.’’ For the third one analysis, the server again used FastTree server, and the neighbour join tree algorithm with a local bootstrap based method on 100 samples to calculate data. It reflected next to the tree nodes. FastTree determine roughly maximum-likelihood for the development of phylogenetic trees from this for family alignment. As B1-subfamily proteins family is not available, we used the secretin family (Family: 7tm_2 (PF00002)) for further analysis. This server also uses multiple sequence alignments and Hidden Markov models (HMMs) algorithm. We developed above three types of trees.

Results and Discussion

The ‘B1 subfamily’ members of the secretin receptor family were retrieved from NCBI, and the information consisting of accession number, location, amino acid sequence length, etc. depicted in Table S1. The proteins of the particular family member and their genes were documented (Table S2). This subfamily contains only 15 protein members, and therefore, this was our sample size; we analysed 13 proteins in our study ($n = 13$) due to the sequence availability. However, several medically necessary small protein families were available, which contained 5 to 15 protein members that were sufficient to analyse for proper scientific understanding. Several studies have been performed in this direction, such as with the insulin receptor

substrate family (Chakraborty et al. 2011) and glucose transporter family (Wood and Trayhurn 2003).

Multiple Sequence Alignment (MSA) and Score Determination

The result of the MSA is documented in Figure S1. Most of the aligned sequences are found between positions 26 to 40; 251 to 299; 350 to 415, 425 to 445, 495 to 515 and 531 to 550. The graphical representation of alignment scores between two proteins is shown in Fig. 1a, b, c. The analysis of sequence alignment shows a maximum score of 67 between the sequence of CRHR1 and CRHR2. The minimum alignment score of 21 was experiential between the sequence of calcitonin hormone receptor 1 and glucagon-like peptide receptor 2. Then, we applied Gblocks to create a graphical view of the extremely associated blocks among the

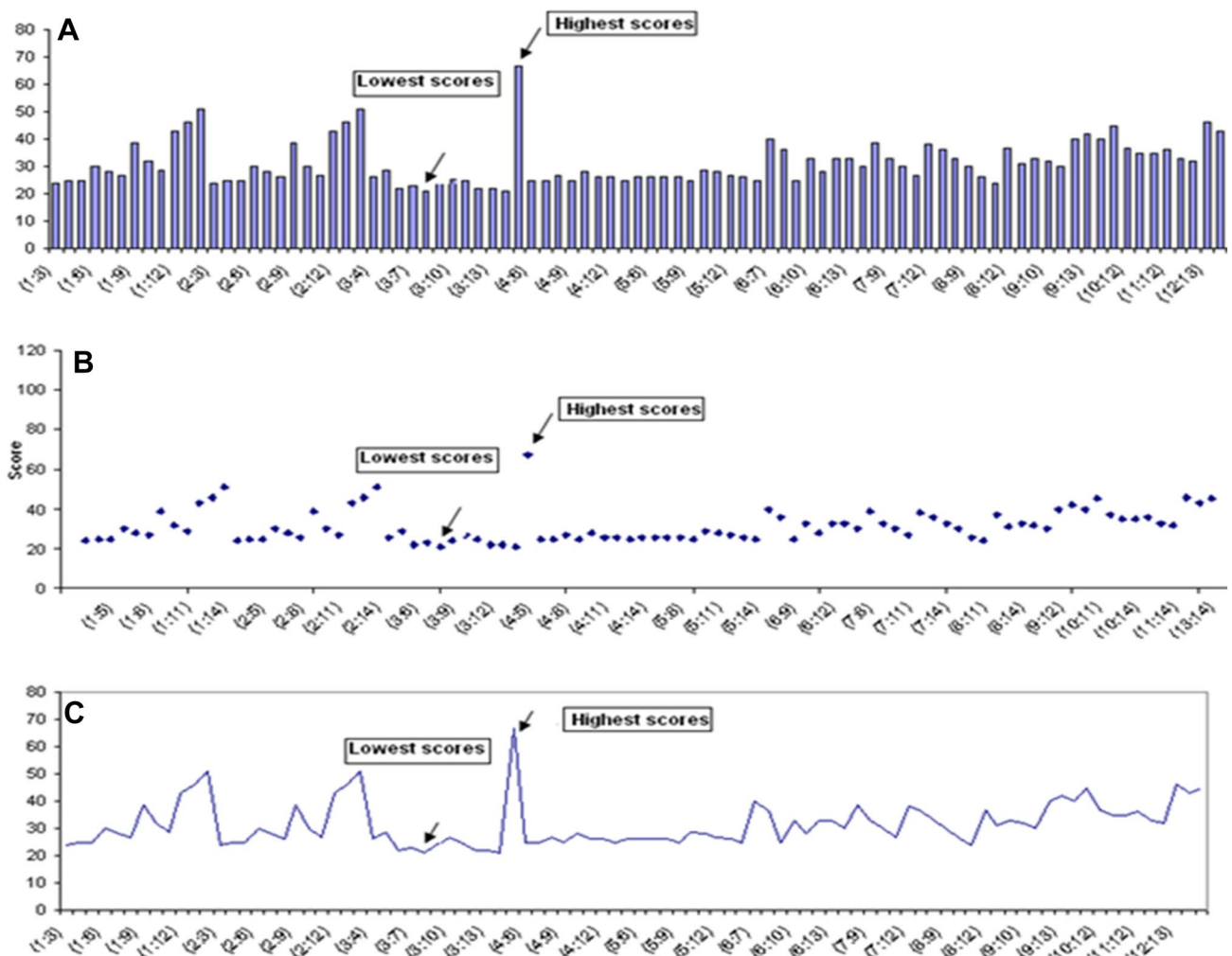


Fig. 1 Alignment scores of protein sequences related to B1- subfamily members. **a** Alignment score between sequences (notation Seq (x:y) meaning alignment score between sequence x, and sequence y);

b Scatter distribution of scores; **c** scores connected by smoothed line without marker

B1-subfamily members (Fig. S2). The highly aligned blocks were visualised between sequence number 254 to 302, 344 to 372, 426 to 455, 494 to 554 and 539 to 564, respectively. Harmar performed multiple sequence alignment of family-B GPCRs, and seven aligned blocks were observed. In the same study, putative hormone-binding domain (PHBD) was also evaluated which is located in the extracellular area of family-B GPCRs (Harmar 2001). However, in our study with B1 subfamily members, we found six multiple sequence aligned blocks and five highly aligned blocks (Gblocks).

Phylogenetic Tree Building as well as Bootstrap Estimation Study

Constructed phylogenetic tree has been shown the useful relatedness among the different sequences of the B1-subfamily members. Then, we developed a phylogram, and the trees that were generated are shown in Fig. 2. The phylogram represents branch length proportional to evolutionary time. In this study, we found that CRHR1 and CRHR2 have a common ancestral relationship and were siblings in 100% bootstrap values; GIPR and GLP1R were siblings in 87% bootstrap values, and VIPR2 and ADCYAP1R1 were siblings in 72% bootstrap values. It must be noted that two receptors, SCTR and GHRHR, exhibited a common point of origin with 29% bootstrap replications. However, SCTR and GHRHR exhibited the lowest bootstrap replications, and CRHR1 and CRHR2 showed the highest bootstrap replications among receptors. The evolutionary relationship between GPCRs was developed by Graul and Sadee (2001) using a clustered database approach, where they used BLAST to yield a database of approximately 1700 GPCRs. However, no report was generated for the phylogenetic relationship between B1-subfamily members. Here,

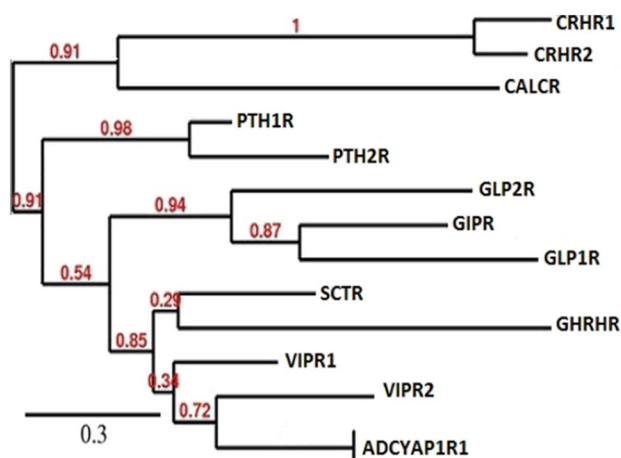


Fig. 2 Phylogenetic tree for B1 subfamily members (construction using Phylogeny.fr software). The Phylogram tree shows bootstrap values at the inner nodes

this phylogenetic analysis of 13 members of the B1-subfamily shows significant evolutionary relationships between members.

Cladogram Analysis Using Computational Biology And Unrooted Tree Formation

A cladogram uses a cladistics method to correlate parental relationships. Hence, cladograms are more informative than phylograms. Figure 3a shows the cladogram of protein sequences of the B1-subfamily members. Figure 3b shows a cladogram with the equal representation of nodes, and binary tree equivalents to the cladogram are also shown. In the matching illustration of nodes, a total of 24 nodes was observed in the cladogram figure (Fig. 3c). In broad-spectrum, the cladogram on ‘n’ number of species (B1-subfamily members) has ‘2n-1’ edges, and the number of search Q(n) for every protein in a particular cladogram tree (phylogenetic tree) is in the variety of $\log n \leq Q(n) \leq n$ (where number of nodes in a binary tree is represented as n) (Deo 2017; Mittal et al. 1994). In the view of the algorithmic point, at the end of level 0, one node is likely. At the point of level 1, two nodes are possible and it will go on like this way. Therefore, the total number of nodes for the binary tree should be $20 + 21 + 22 + \dots + 2p \geq n$ (at p level). The distance of the path between two leaf nodes established the relationship in a binary tree. This binary tree is a six-level binary tree, and VIPR2 and ADCYAP1R1 are situated in the leaf node. Graphical representation was provided between the number of nodes in every level (Fig. 3d), where it must be noted that level-3 and level-5 obtained a maximum number of six nodes.

Two types of unrooted trees were illustrated to understand the possible relationships between B1-subfamily members, the unrooted-alpha tree and the circular alpha tree. Both trees are shown in Fig. 4. Nine internal nodes and thirteen external nodes were demonstrated in the unrooted-alpha tree. Conversely, eleven internal nodes and thirteen external nodes were illustrated in the circular alpha tree. John et al. (2003) stated that unrooted binary trees are also crucial for the conclusion of evolutionary trees based on data condition for all leaf node species and the unrooted binary tree describing the evolution of those species. However, in this case, our tree depicts branch-decomposition of these proteins, and these unrooted trees can solve the natural graph enumeration problem (Balding et al. 2008).

Conservation Patterns Examination and Computation of Extremely Conserved Amino Acids

The conservation prototype and protein backbone structures with extremely conserved parts and the number of extremely

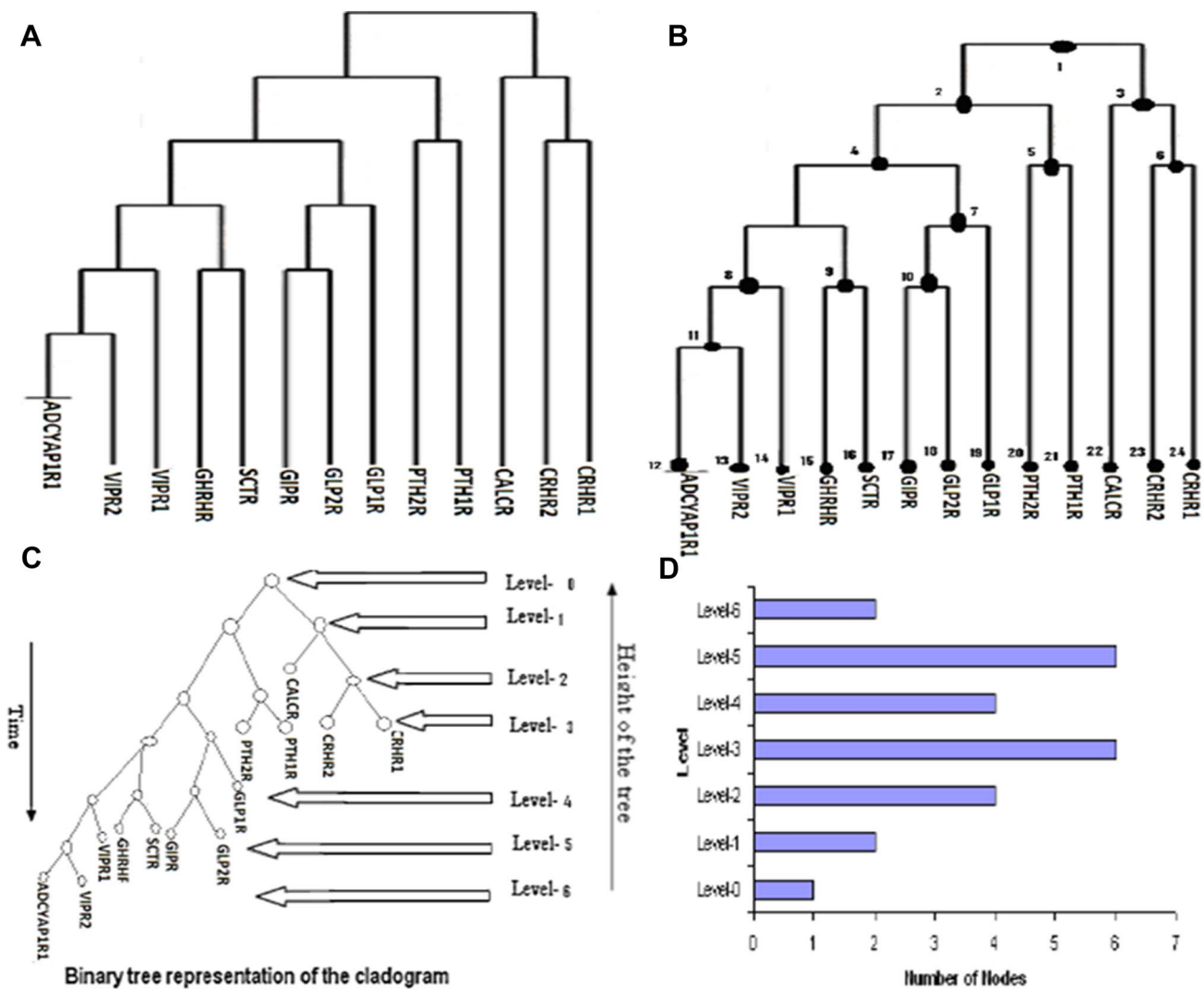


Fig. 3 Cladogram (B1 subfamily proteins) and its analysis using computational biology. **a** Cladogram for tree algorithm analysis **b** Matching representation of nodes **c** Representation binary tree equivalent to Cladogram **d** Number of nodes in each level

conserved protein residues are shown in Fig. 5. CRHR2 has the maximum amount of extremely conserved amino acid residues followed by the PACP1 receptor. VIPR2 contained the lowest number of highly conserved residues, whereas the calcitonin receptor, secretin receptor and VIPR1 lacked highly conserved amino acid residues. Some typical features were observed in B1-subfamily members. The most common semi-conserved structural characteristics of the B1-family GPCR are a comparatively small C-terminus and an elongated N-terminal domain with a residue of 100–200 aa (amino acids) (George et al. 2002).

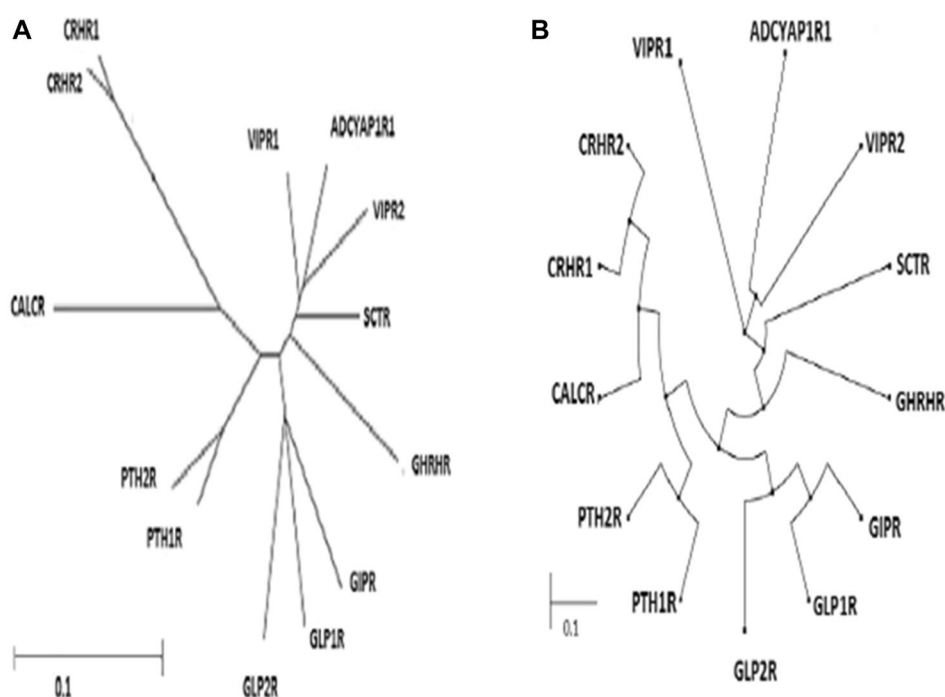
Moreover, in the extracellular loop (ECL) ECL 1 and ECL2, a conserved disulphide bridge was noted between cysteine residues (Authier and Desbuquois 2008; Kenakin and Miller 2010; Ulrich et al. 1998b). Particularly, conserved regions contribute in biochemical performance such

as binding regions of surface structure on the diverse exterior part of proteins (Ghosh et al. 2012); (Branden and Tooze 1999); (Ulrich et al. 1998a). The highly conserved regions of membrane proteins or viral peptides can be used for vaccine development, and we have found several conserved residues within B1-subfamily members in this study.

Sequence Logos of Conserved Residues

Sequence logos of conserved residues are depicted in Fig. 6. The position of every single amino acid is represented in the form of its unique symbol; the height of the symbol pointed out the virtual frequency of that particular amino acid at an exact position in the sequence. We found sequence logos in several continuous positions, including 6 to 21, 27 to 43, 70 to 78, 85 to 94, 142 to

Fig. 4 The unrooted phylogenetic tree of B1 subfamily proteins. **a** Unrooted-alpha phylogenetic tree **b** Circular alpha phylogenetic tree



153, 155 to 160, 162 to 181, 205 to 220, 227 to 238, 240 to 248, 252 to 312, 316 to 324, 349 to 423, 425 to 430, 432 to 472, 404 to 430, 434 to 574, 583 to 585, 603 to 605, 609 630, 646 to 648, 652 to 654, and 674 to 678. The longest continuous number sequence logo (74) was found between sequence positions 349 to 423. In this study, the maximum height recorded logo is 3.6 bits, and the minimum height recorded is 0.18 bits. This enabled us to visually inspect and examine the significant amino acids that constitute a functionally significant area of the protein. Perrin et al. (2007) found that human B1 family of GPCR show conservation of structurally essential residues comprise of the salt bridge, six cysteines residues, two tryptophan residues, a proline and a glycine. Cysteine residues are required for conserved disulphide bridge formation, and this was confirmed from in our study. However, structure–function relationships proposed by Perrin et al. (2007) recognized a structurally significant salt bridge and aspartic acid residue which is extremely conserved. It is located in the initial segment extracellular domain (ECD1) of receptors in B1 GPCR family members, which are important for the functionality and structure formation of the ‘family B1’ of this group of receptors. This receptor class is significant for drug development. Therefore, the protein domains that contain these conserved residues are highly important and can be used for drug development, drug targeting, or drug interaction studies (Giaccia et al. 2003). We can understand protein–protein interactions from these highly conserved segments (Li et al. 1998).

Distribution of B1-Subfamily Proteins Among Different Species

A graphical representation of the distribution of secretin family member proteins (which contains B1-subfamily receptor proteins) across species is shown in Fig. 7. The tree is built by considering the taxonomic lineage of each sequence that has a match to these family members. Three types of trees, a “sunburst” tree (Fig. 7a), a “species tree” (Fig. 7b), and a “phylogenetic tree for family seed alignment” (Fig. 7c), are depicted. Two types of “sunburst” trees were developed: tree segments weight by the number of sequences and tree segments weight by the number of species. These “sunburst” trees show each node in the tree as a separate arc, arranged radially with the super kingdoms at the centre and the species arrayed around the outermost ring. This tree information indicates that secretin family member proteins are not only distributed in *Homo sapiens* but also distributed in several species, including Insecta, Diptera, Cyprinidae, and Mammalia. The “species tree” shows the occurrence of this subfamily across different species, including 247 species of Eukaryota, 190 Metazoa and two Choanoflagellida. Harmar et al. concluded that the subfamily was encoded by 15 genes in humans, as well as at least five genes in *Drosophila* and three in *C. elegans* (Harmar 2001).

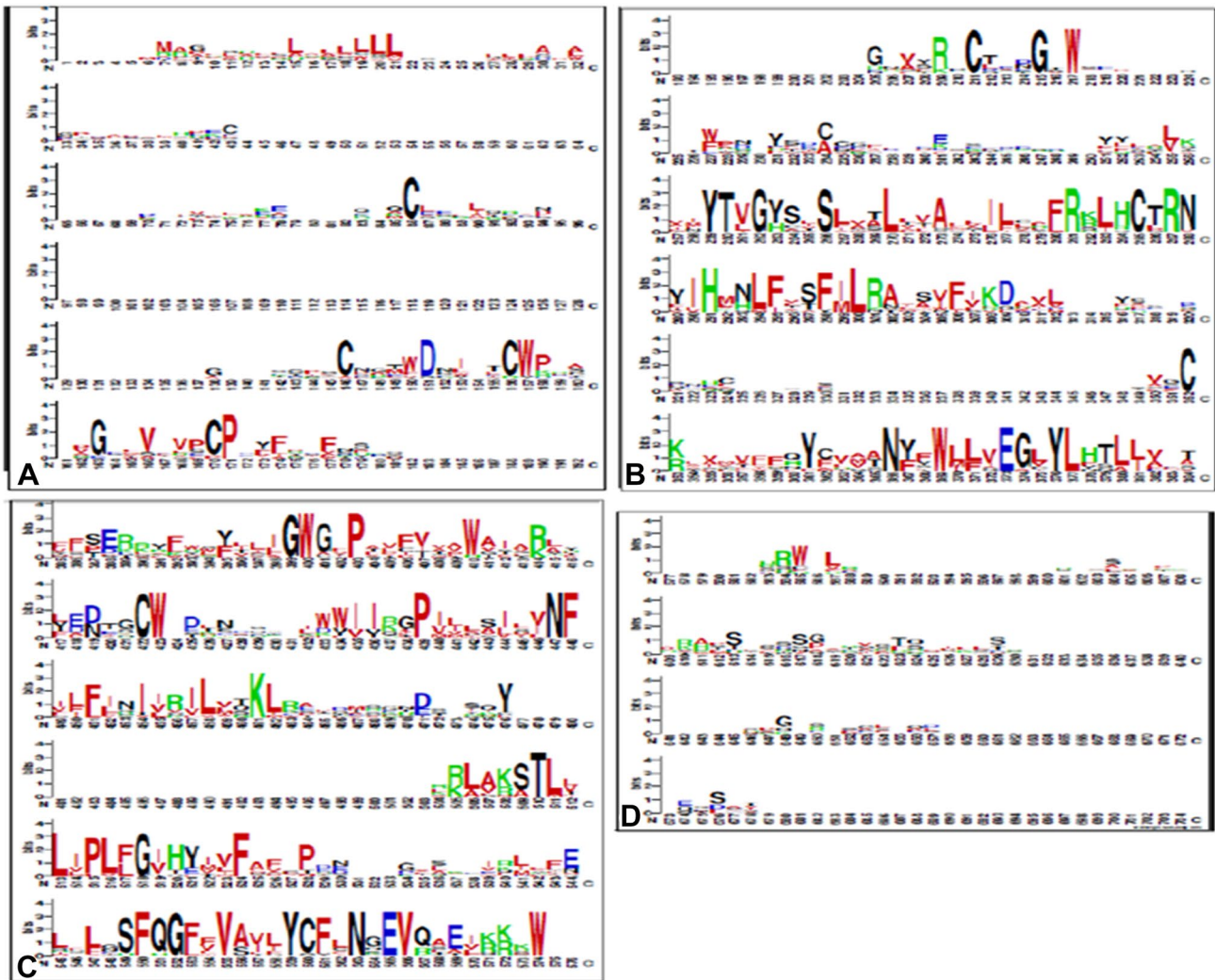


Fig. 6 Generated WebLogo for B1 subfamily proteins in different sequence position **a** 1 to 192 sequence **b** 193 to 384 sequence **c** 385 to 576 sequences **d** 577 to 704 sequences

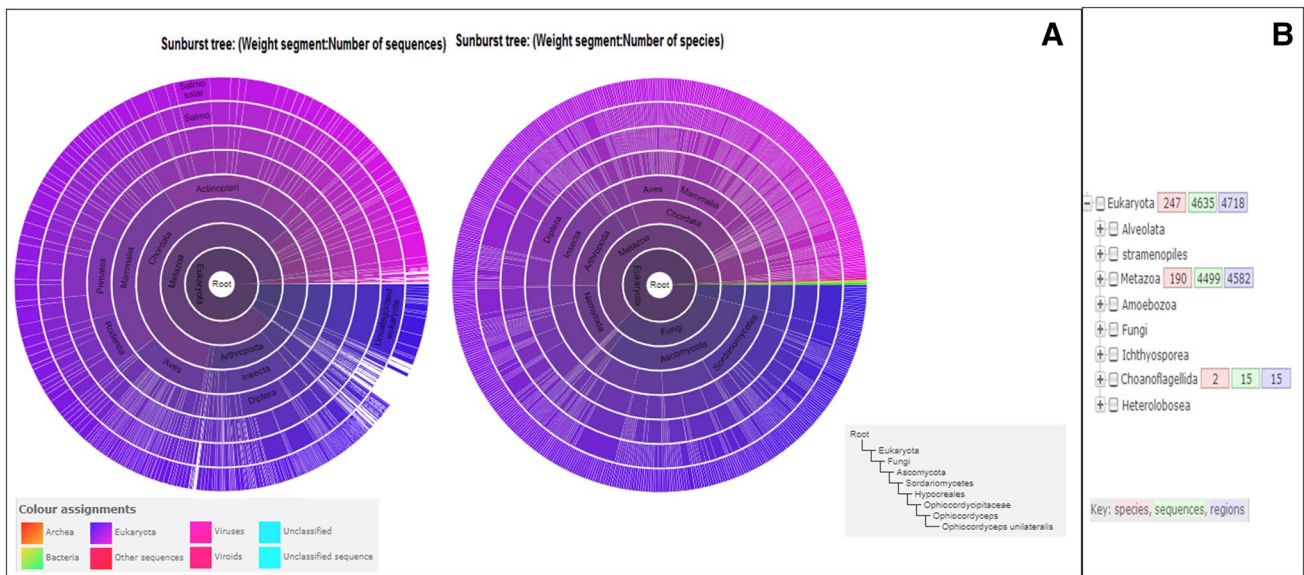


Fig. 7 Graphical representation of the distribution of secretin receptor family proteins across species. **a** "sunburst" visualization of the species tree for this family. It shows each node in the tree as a separate arc, arranged radially with the super kingdoms at the centre and the species arrayed around the outermost ring. **b** The "species tree"

which shows the occurrence of this subfamily across different species. **c** The "phylogenetic tree for secretin family seed alignment" which uses neighbor join tree algorithm with a local bootstrap based on 100 resamples (which is reflected next to the tree nodes)

Acknowledgements This research was supported by Hallym University Research Fund, and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1A2B4012944).

Research Involving Human and Animal Rights This article does not contain any studies with human participants or animals performed by any of the authors.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Aguilera G, Wynn PC, Harwood JP, Hauger RL, Millan MA, Grewe C, Catt KJ (1986) Receptor-mediated actions of

- corticotropin-releasing factor in pituitary gland and nervous system. *Neuroendocrinology* 43:79–88
- Alpini G, Ulrich C, Phillips J, Pham L, Miller LJ, LaRusso NF (1994) Upregulation of secretin receptor gene expression in rat cholangiocytes after bile duct ligation. *Am J Physiol-Gastrointest Liver Physiol* 266(5):G922–G928
- Andersson A, Sundler F, Ekblad E (2000) Expression and motor effects of secretin in small and large intestine of the rat. *Peptides* 21(11):1687–1694
- Araç D, Leon K (2020) Structure, function and therapeutic potential of adhesion GPCRs. *GPCRs*. Academic Press, Cambridge, pp 23–41
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38(2):W529–W533
- Atkinson GC, Haurlyuk V, Tenson T (2011) An ancient family of SelB elongation factor-like proteins with a broad but disjunct distribution across archaea. *BMC Evol Biol* 11:22
- Authier F, Desbuquois B (2008) Glucagon receptors. *Cell Mol Life Sci* 65(12):1880–1899
- Balding DJ, Bishop M, Cannings C (2008) Handbook of statistical genetics. John Wiley & Sons, New Jersey
- Bawab W, Gaspach C, Marie J-C, Chastre E, Rosselin G (1988) Pharmacology and molecular identification of secretin receptors in rat gastric glands. *Life Sci* 42:791–798
- Bhattacharya P, Yan YL, Postlethwait J, Rubin DA (2011) Evolution of the vertebrate pth2 (tip39) gene family and the regulation of PTH type 2 receptor (pth2r) and its endogenous ligand pth2 by hedgehog signaling in zebrafish development. *J Endocrinol* 211(2):187–200
- Branden CI (1999) Introduction to protein structure. CRC Press, Boca Raton, Garland Science
- Calvi L, Schipani E (2000) The PTH/PTHrP receptor in Jansen's metaphyseal chondrodysplasia. *J Endocrinol Invest* 23:545–554
- Cardoso J, Clark M, Viera F, Bridge P, Gilles A, Power D (2005) The secretin G-protein-coupled receptor family: teleost receptors. *J Mol Endocrinol* 34:753–765
- Cardoso J, Félix R, Power DM (2014) Nematode and arthropod genomes provide new insights into the evolution of class 2 B1 GPCRs. *PLoS ONE* 9(3):e92220
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552
- Chakraborty C, Agoramoorthy G (2014) Comparative bioinformatic analysis of the conserved domains, amino acid residues, and binding grooves of tumor necrosis factor. *Med Glas* 11(1):1–6
- Chakraborty C, Agoramoorthy G, Hsu MJ (2011) Exploring the evolutionary relationship of insulin receptor substrate family using computational biology. *PLoS ONE* 6:e16580
- Chambers TJ, Magnus CJ (1982) Calcitonin alters behaviour of isolated osteoclasts. *J Pathol* 136:27–39. <https://doi.org/10.1002/path.1711360104>
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500
- Chow B, Cheung K, Tsang E, Leung M, Lee S, Wong P (2004) Secretin controls anion secretion in the rat epididymis in an autocrine/paracrine fashion. *Biol Reprod* 70:1594–1599
- Christophe J, Chatelain P, Taton G, Delhay M, Waelbroeck M, Robberecht P (1981) Comparison of VIP-secretin receptors in rat and human lung. *Peptides* 2:253–258
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190
- Dacquin R et al (2004) Amylin inhibits bone resorption while the calcitonin receptor controls bone formation in vivo. *J Cell Biol* 164:509–514
- Davey RA et al (2008) Calcitonin receptor plays a physiological role to protect against hypercalcemia in mice. *J Bone Miner Res* 23:1182–1193
- Dayhoff M (1969) Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington DC
- Deo N (2017) Graph theory with applications to engineering and computer science. Courier Dover Publications, USA
- Dereeper A et al (2008) Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:W465–W469
- Dong M, Miller LJ (2002) Molecular pharmacology of the secretin receptor. *Recept Channels* 8:189–200
- Drucker DJ, Philippe J, Mojsov S, Chick WL, Habener JF (1987) Glucagon-like peptide I stimulates insulin gene expression and increases cyclic AMP levels in a rat islet cell line. *Proc Natl Acad Sci* 84:3434–3438
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Felley C, Qian J, Mantey S, Pradhan T, Jensen R (1992) Chief cells possess a receptor with high affinity for PACAP and VIP that stimulates pepsinogen release. *Am J Physiol-Gastrointest Liver Physiol* 263(6):G901–G907
- Fremeau RT, Jensen RT, Charlton CG, Miller RL, O'Donohue TL, Moody TW (1983) Secretin: specific binding to rat brain membranes. *J Neurosci* 3(8):1620–1625
- George SR, O'Dowd BF, Lee SP (2002) G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nat Rev Drug Discov* 1:808–820
- Gaspach C, Bataille D, Vauclin N, Rosselin G, Moroder L, Wunsch E (1981) Secretin binding sites coupled with adenylate cyclase in rat fundic membranes. *Peptides* 2:247–251
- Ghosh A, Chattopadhyay S, Chawla-Sarkar M, Nandy P, Nandy A (2012) In silico study of rotavirus VP7 surface accessible conserved regions for antiviral drug/vaccine design. *PLoS ONE* 7(7):e40749
- Giaccia A, Siim BG, Johnson RS (2003) HIF-1 as a target for drug development. *Nat Rev Drug Discov* 2(10):803–811
- Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 19(1):163–164
- Grammatopoulos DK, Chrousos GP (2002) Functional characteristics of CRH receptors and potential clinical applications of CRH-receptor antagonists. *Trends Endocrinol Metab* 13:436–444
- Graul RC, Sadée W (2001) Evolutionary relationships among G protein-coupled receptors using a clustered database approach. *AAPS J* 3:25–42
- Hamby SE, Hirst JD (2008) Prediction of glycosylation sites using random forests. *BMC Bioinf* 9(1):500
- Harmar AJ (2001) Family-B G-protein-coupled receptors. *Genome Biol* 2(12):3013
- Hauger RL, Grigoriadis DE, Dallman MF, Plotsky PM, Vale WW, Dautzenberg FM (2003) International Union of Pharmacology. XXXVI. Current status of the nomenclature for receptors for corticotropin-releasing factor and their ligands. *Pharmacol Rev* 55(1):21–26
- Herrada EA, Tessone CJ, Klemm K, Eguíluz VM, Hernández-García E, Duarte CM (2008) Universal scaling in the branching of the tree of life. *PLoS ONE* 3:e2757
- Herrada A, Eguíluz VM, Hernández-García E, Duarte CM (2011) Scaling properties of protein family phylogenies. *BMC Evol Biol* 11:155

- Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 42:182–192
- Holmes S (2003) Bootstrapping phylogenetic trees: theory and methods. *Stat Sci* 18(2):241–255
- Ishihara T, Nakamura S, Kaziro Y, Takahashi T, Takahashi K, Nagata S (1991) Molecular cloning and expression of a cDNA encoding the secretin receptor. *EMBO J* 10:1635
- John KS, Warnow T, Moret BM, Vawter L (2003) Performance study of phylogenetic methods:(unweighted) quartet methods and neighbor-joining. *J Algorithms* 48(1):173–193
- Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298
- Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: the impact of allosteric modulation and functional selectivity on new drug discovery. *Pharmacol Rev* 62:265–304
- Kinzig KP, D'Alessio DA, Seeley RJ (2002) The diverse roles of specific GLP-1 receptors in the control of food intake and the response to visceral illness. *J Neurosci* 22(23):10470–10476
- Kolakowski LF Jr (1994) GCRDb: a G-protein-coupled receptor database. *Recept Channels* 2(1):1–7
- Köves K, Kausz M, Reser D, Horvath K (2002) What may be the anatomical basis that secretin can improve the mental functions in autism? *Regul Pept* 109:167–172
- Laburthe M, Couvineau A (2002) Molecular pharmacology and structure of VPAC receptors for VIP and PACAP. *Regul Pept* 108:165–173
- Letunic I, Bork P (2006) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1):127–128
- Li P, Chang T, Chey W (1998) Secretin inhibits gastric acid secretion via a vagal afferent pathway in rats. *Am J Physiol-Gastrointest Liver Physiol* 275(1):G22–G28
- Mayrose I, Graur D, Ben-Tal N, Pupko T (2004) Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol Biol Evol* 21:1781–1791
- Mittal R, Jain BN, Patney RK (1994) Finding nonfaulty subtrees in faulty binary tree architectures. *Microelectron Reliab* 34:1301–1310
- Murray RA, Maheshwari HG, Russell EJ, Baumann G (2000) Pituitary hypoplasia in patients with a mutation in the growth hormone-releasing hormone receptor gene. *Am J Neuroradiol* 21(4):685–689
- Nozaki S, Nakata R, Mizuma H, Nishimura N, Watanabe Y, Kohashi R, Watanabe Y (2002) In vitro autoradiographic localization of 125I-secretin receptor binding sites in rat brain. *Biochem Biophys Res Commun* 292(1):133–137
- Ogi K et al (1993) Molecular cloning and functional expression of a cDNA encoding a human pituitary adenylate cyclase activating polypeptide receptor. *Biochem Biophys Res Commun* 196:1511–1521
- O'hayre M, Vázquez-Prado J, Kufareva I, Stawiski EW, Handel TM, Seshagiri S, Gutkind JS (2013) The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. *Nat Rev Cancer* 13(6):412–424
- Peng X et al (2012) CCCH-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. *PLoS ONE* 7(7):e40120
- Perrin MH et al (2007) Distinct structural and functional roles of conserved residues in the first extracellular domain of receptors for corticotropin-releasing factor and related G-protein-coupled receptors. *J Biol Chem* 282:37529–37536
- Punta M et al (2011) The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301
- Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* 18(1):S71–S77
- Purcell R, Hall RA (2018) Adhesion G protein-coupled receptors as drug targets. *Annu Rev Pharmacol Toxicol* 58:429–449
- Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18:6097–6100
- Siu FK, Lam IP, Chu JY, Chow BK (2006) Signaling mechanisms of secretin receptor. *Regul Pept* 137(1–2):95–104
- Stoffel M, Fernald AA, Le Beau MM, Bell GI (1995) Assignment of the gastric inhibitory polypeptide receptor gene (GIPR) to chromosome bands 19q13.2–q13.3 by fluorescence in situ hybridization. *Genomics* 28(3):607–609
- Ulrich CD, Holtmann M, Miller LJ (1998a) Secretin and vasoactive intestinal peptide receptors: Members of a unique family of G protein-coupled receptors. *Gastroenterology* 114(2):382–397
- Ulrich CD, Wood P, Hadac EM, Kopras E, Whitcomb DC, Miller LJ (1998b) Cellular distribution of secretin receptor expression in rat pancreas. *Am J Physiol-Gastrointest Liver Physiol* 275(6):G1437–G1444
- Vass M, Kooistra A, Yang D, Stevens RC, Wang MW, de Graaf C (2018) Chemical diversity in the G protein-coupled receptor superfamily. *Trends Pharmacol Sci* 39(5):494–512
- Vaudry D, Gonzalez BJ, Basille M, Yon L, Fournier A, Vaudry H (2000) Pituitary adenylate cyclase-activating polypeptide and its receptors: from structure to functions. *Pharmacol Rev* 52:269–324
- Venkatakrishnan A, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM (2013) Molecular signatures of G-protein-coupled receptors. *Nature* 494(7436):185–194
- Wheeler DL et al (2007) Database resources of the national center for biotechnology information. *Nucleic Acids Res* 36:D13–D21
- Wood IS, Trayhurn P (2003) Glucose transporters (GLUT and SGLT): expanded families of sugar transport proteins. *Br J Nutr* 89(1):3–9
- Yung W-H, Leung P-S, Ng SS, Zhang J, Chan SC, Chow BK (2001) Secretin facilitates GABA transmission in the cerebellum. *J Neurosci* 21:7063–7068

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.