Journal of Peptide Science
J. Pept. Sci. 2007; **13**: 568–574
Published online 13 August 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.898

JOURNAL OF PEPTIDE SCIENCE

Bioinformatics-based discovery and identification of new biologically active peptides for GPCR deorphanization‡

JEAN COLETTE, EMILIE AVÉ, BENJAMIN GRENIER-BOLEY, ANNE-SOPHIE COQUEL, KARINE LESELLIER and KARINE PUGET*
Genepep S.A., Zone d'activité les Baronnes, Ateliers Relais, Lots 4 & 5, 34 730 Prades le Lez, France

Received 29 March 2007; Accepted 4 June 2007

Abstract: Owing to their involvement in many physiological and pathological processes, G-protein-coupled receptors (GPCRs) are interesting targets for drug development. Approximately, 100 endoGPCRs lack their natural ligands and remain orphan (oGPCRs). Consequently, oGPCR deorphanization appears a promising research field for the development of new therapeutics. On the basis of the knowledge of currently known GPCR/ligand couples, some oGPCRs may be targeted by peptides. However, to find new drugs for GPCRs, Genepep has developed a dedicated bioinformatics platform to screen transcriptomic databases for the prediction of new GPCR ligands. The peptide lists generated include specific data, such as chemical and physical properties, the occurrence of post-translational modifications (PTMs) and an annotation referring to the location and expression level of the related putative genes. This information system allows a selection through series of biological criteria of ∼10 000 natural peptides including already known GPCR ligands and potential new candidates for GPCR deorphanization. The most promising peptides for functional assay screening and future development as therapeutic agents are under evaluation. Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: GPCR; deorphanization; bioinformatics; drug discovery; peptide; SAGE; oncology

INTRODUCTION

The G-protein-coupled receptors (GPCRs) form the largest class of cell surface receptors and regulate various cellular functions responsible for physiological responses. GPCRs represent one of the major targets for modern pharmaceutical drugs. Over 50% approved drugs elicit their therapeutic effects on selected members of the GPCRs [1]. With nearly 800 members, the GPCR superfamily represents the largest human gene family [2]. On the basis of their specific ligand-coupling pattern, the GPCR superfamily can be divided into GPCRs with sensory signals of external origin (chemosensory GPCRs or csGPCRs) and GPCRs with endogenous ligands (endoGPCRs). The csGPCRs are mainly targeted by olfactory/gustatory ligands (chemokines and chemoattractants), while the endoGPCRs are targeted by a wide variety of ligands (opsins, biogenic amines, lipid mediators and ligand peptides) [3]. Among the 367 endoGPCRs listed in the human genome [2], 140 are still GPCRs with no identified natural ligand and function (referred as orphan GPCRs, oGPCRs) [3]. In 2001, Civelli estimated that at least 50 natural peptidic ligands remain to be discovered [3]. Indeed, among the human endoGPCRs, more than a third are targeted by peptides. Consequently, finding natural or synthetic derived ligands that activate oGPCR appears as a promising research field for the identification and development of new peptidic therapeutics.

Since the first reports on oGPCR deorphanizations [4,5], intensive research efforts based on reverse pharmacology [6] have led to some deorphanization. The pairing of known ligands [7], the use of tissue extracts [3] and the random high-throughput screening (HTS) of large libraries of synthetic or natural molecules [8] have successively allowed the discovery of over 100 new ligands [3,9]. Surprisingly, since the mid-1990s, only a dozen of novel ligands have been discovered [6,9]. The advent of the genomic era and the emergence of bioinformatics have opened new insights in the field of oGPCR deorphanization. The *in silico* prediction of novel peptidic ligands for GPCRs by screening the human transcriptome databases has already shown its efficiency with the identification of the QRFP/P52 peptide. This peptide, a new member of the RFamide family, has been predicted *in silico* by searching a *C*-terminal RFG[KR] motif in a virtual protein transcript database (VTS). QRFP/P52 has been shown to selectively activate the oGPCR SP9155/GPR103 [10]. However, the bioinformatics screening of human transcriptome databases with a pre-determined motif remains restrictive for different reasons. First, the primary structures of GPCR peptidic ligands are highly diverse, and no clear-cut consensus sequence can be evidenced. Second, the use of a consensus motif defined

*Correspondence to: Karine Puget, Genepep S.A., Zone d'activité les Baronnes, Ateliers Relais, Lots 4 & 5, 34 730 Prades le Lez, France; e-mail: puget@genepep.com
‡ This article is published as part of the special issue 2nd International Congress on Natural Peptides to Drugs, April 18–21, 2006, Zermatt, Switzerland.

Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

WILEY InterScience® DISCOVER SOMETHING GREAT

from a multiple sequence alignment could result in identifying peptides with ligand properties identical to the ones observed for the peptides used for generating the consensus motif. Moreover, many GPCRs are highly conserved among species. Indeed, all the major GPCR families found in the human genome were shown to arise before the evolutionary split of the nematodes from the chordate lineage [11]. This suggests that peptidic ligands would also be conserved from species to species and would activate GPCRs beyond species boundaries. This has been already documented with the example of bombesin. Bombesin, which is a 14-residue peptide isolated from the skin of amphibian frogs [12], was shown, among different pharmacological effects, to increase insulin secretion and to stimulate uterus contraction in mammals [13]. Subsequently, the bombesin-like gastrin-releasing peptides (GRPs) and neuromedin B (NMB) have been isolated from porcine [14,15] and human tissues [16,17]. Thus, the bioinformatics screening of non-human genomes would be an added value for the discovery of novel peptidic ligands. In the frame of the discovery of new ligands, knowledge of their bearing post-translational modifications (PTMs) represents also a key issue. In fact, PTMs may alter physical and chemical properties, folding, conformation distribution, stability, activity and function of the peptides. For example, in the cholecystokinin (CCK), the sulfation of the tyrosine in position 27 is critical for affinity and potency of a full agonist of the type-A CCK receptor [18]. In addition, many bioactive peptides are C-terminally amidated. Such a modification appears to be involved in biological activity [19], to be part of the peptide stability [20], and in some cases to play a significant role in the bioactive conformation of the peptide or its interaction with a receptor [21]. Consequently, prediction of PTMs is an important issue in a search procedure for a new therapeutic with ligand properties.

Finally, the bioinformatics screening of animal and plant genomes is likely to provide long lists of new peptidic ligands. Indeed, genome bioinformatics screenings generally use prediction methods. These methods are based on motif recognition and/or machine learning methods applied to primary sequences and physico-chemical or structural characteristics of peptides and proteins. As any prediction method can generate false-positive predictions, combining the data obtained through the genomic screening to other specific data such as expression patterns would allow selecting the predicted peptides with the highest probability to represent the target of choice. Indeed, genes encoding the GPCR peptidic ligand can show specific expression profiles. For instance, the human gastrin gene is specifically expressed in G-cells of the stomach [22], while the human adrenomedullin gene is highly expressed in glioblastoma cell lines [23]. The processing of the data obtained by the serious analysis of gene expression (SAGE) method [24] applied to various tissue samples could lead to the determination of the location and expression levels in different physiological and pathological contexts (tumoral *vs* normal, infected *vs* healthy) of the putative genes. On the basis of these observations, Genepep has set up a bioinformatics platform to screen transcriptomic and protein databases to predict novel GPCR peptidic ligands, their PTMs and their local gene expression levels thanks to the SAGE databases. This platform is based on a series of predictive programs simulating the natural process of a peptide precursor synthesis and maturation (translation, enzymatic cleavages, PTMs). These predictive programs are based on machine learning approaches trained with biological data about known GPCR peptidic ligands. Our proprietary-owned bioinformatics platform also includes programs that allow the annotation of the generated peptides by annotation transfer from the original genome sequences and the screening of expression data coming from a SAGE library database. The best GPCR peptidic ligand candidates can finally be selected thanks to a user-friendly interface in which one can mention various selection criteria (species, organisms, tissue expression, amino acid composition and molecular mass, etc.).

STRATEGY AND BIOINFORMATICS PLATFORM

General Strategy

Our strategy for the discovery of novel GPCR peptidic ligand is based on a multi-step process (Figure 1). Briefly, all the available eukaryotic genome (excepting the plant genomes) databases are processed through a Genepep proprietary bioinformatics platform based on two predictive programs: GENE2PEP and PREDILIGAND. The combination of these two programs generates lists of putative GPCR ligand peptides encoded by the screened transcriptomes. The information retrieved includes chemical and physical properties, occurrence of PTMs and location and expression level of the related putative genes. The selected peptides are then synthesized in-house by a solid-phase peptide synthesis at our core facility as libraries of pure peptides prepared in 96-well plates [25]. Finally, the novel GPCR peptidic ligand candidates are used in conventional GPCR functional assays [26].

The Bioinformatics Platform: GENE2PEP & PREDILIGAND

Transcriptomic- and protein-database-related flat files are processed through the GENE2PEP program. GENE2PEP mimics the different natural steps in the synthesis (transcription) and natural processing (enzymatic cleavages, PTM) of peptide precursors. This program generates new and already known peptide

Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

DOI: 10.1002/psc

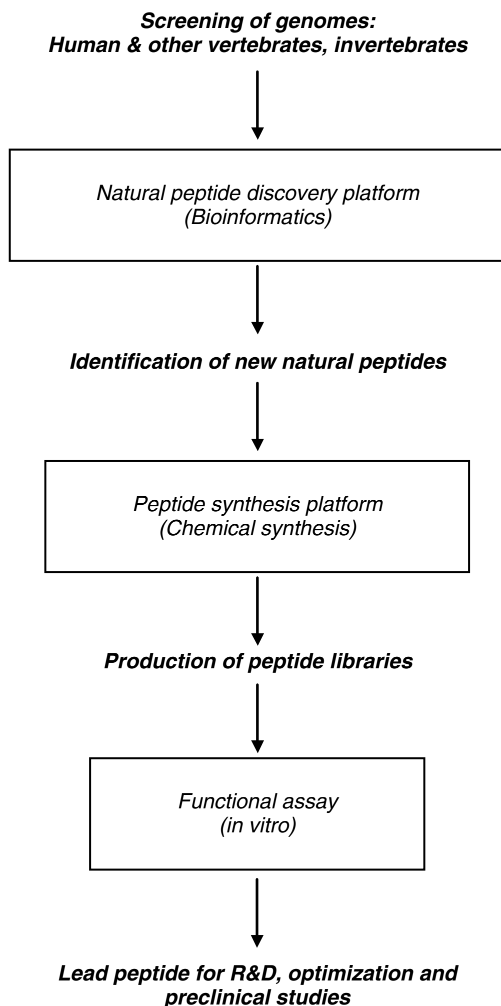

Figure 1 Bioinformatics strategies for novel oGPCRs peptidic ligand discovery.

sequences including their PTMs and annotations coming from related screened sequences.

The screening of the dbEST (expressed sequence tags) database, a division of Genbank [27], and the UniProtKB/Swiss-Prot database [28] is currently supported by GENE2PEP. The dbEST database represents the major source of new sequence records and gene sequences. In early 2007, Genbank release 155 comprised 38.3 million sequences from more than 1200 different organisms [29]. The peptide/protein records of UniProtKB/Swiss-Prot section are the result of full manual and computer assisted/manually controlled annotations based on published literature and sequence analysis [26]. In the UniProt release version 9.0 (October, 2006), the UniProtKB/Swiss-Prot section contained ∼240 000 records (http://expasy.org/sprot/sp-orel.html). Thus, GENE2PEP supports the screening of dbEST and UniProtKB/Swiss-Prot in the form of complete entry flat files that can be retrieved by file transfer protocol (FTP) from the National Center for Biotechnology Information (NCBI) FTP site (ftp.ncbi.nih.gov) and the Expasy FTP server (ftp.expasy.org). From these data, GENE2PEP mimics the natural processing of a precursor polypeptide. *In vivo*, regulatory peptides are initially synthesized as large protein precursors that undergo proteolytic processing to yield the biologically active peptide(s). The processing of the precursor starts by the removal of the signal peptide by an enzyme belonging to the signal peptidase family [30]. Then, proprotein convertase enzymes will generate the bioactive peptide(s). So far, seven proprotein convertase enzymes have been identified in mammals (for a review, see Ref. 31). Convertase enzymes cleave proproteins at single and/or pairs of basic residues (arginine, lysine). However, many precursors are also cleaved at non-basic sites. For instance, cleavages can occur *C*-terminally to alanine, serine, threonine, methionine, valine or leucine [31]. To date, only three enzymes (SKI-1/S1P [32,33], PCSK9/NARC-1 [34] and ECE-2 [35]) with these cleavage specificities have been identified. Additional maturations may also occur at both extremities through the action of exopeptidases (aminopeptidases, dipeptidylpeptidases and/or carboxypeptidases) [36]. Finally, the matured peptides can be subjected to PTMs. GENE2PEP simulates each step of peptide processing by combining first public programs and second Genepep proprietary prediction programs based on machine learning approaches applied on biological sequence data of known natural peptide precursors. The transcription/translation step is performed by the European Molecular Biology Open Software Suite Getorf program (EMBOSS) [37]. The prediction of the cleavage site for signal peptide removal is performed using the SignalP program [38]. Following this two-step procedure, which relies on two public programs, the secreted protein sequences retrieved are processed through two proprietary-owned bioinformatics prediction programs, G2Pcleavage and G2Pptm. G2Pcleavage simulates the cleavages at basic residues and on the exopeptidase-mediated basic residue removal. G2Pptm predicts the occurrence of a selected list of PTMs (amidation, sulfation and cysteine pairing). In order to avoid false sequences resulting from the insertion in the ESTs of segments derived from vectors, GENE2PEP integrates the NCBI Vecscreen program whose function is to identify vector contamination segments (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html). Following this cascade of bioinformatics events, the peptide sequences extracted and processed are submitted to our proprietary-owned PREDILIGAND program.

PREDILIGAND program was set up to identify GPCR-ligand-like peptides among the peptide lists generated by GENE2PEP. This identification is performed by comparing GENE2PEP's predicted peptides *versus* a physico-chemical model of known natural GPCR ligand

Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

DOI: 10.1002/psc

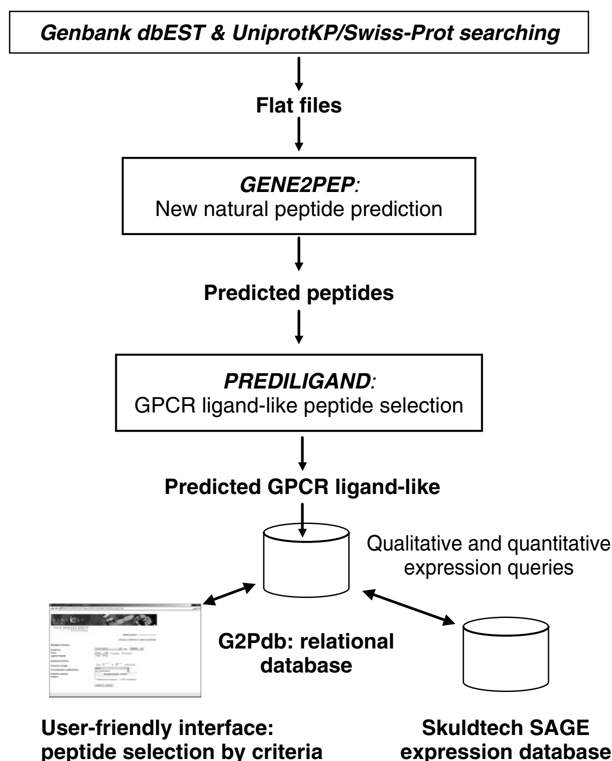

Figure 2 Bioinformatics pipeline for novel oGPCRs peptidic ligand prediction and selection.

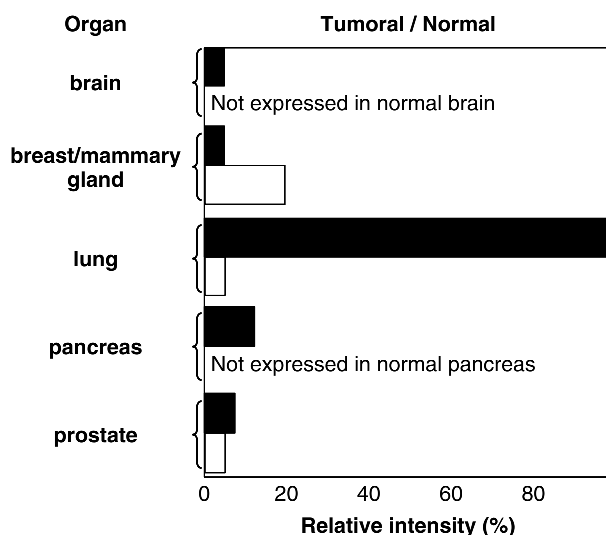

Figure 3 Differential expression profiles using SAGE technology. Normal and tumoral repartition of the 10-nucleotide tag of a putative oGPCR peptidic ligand encoding gene. For each tissue/pathological state couple, the sum of SAGE tag X occurrences is determined. The higher sum is set at 100% and other sums are calibrated between 0 and 100%.

peptides. This model was built using the physico-chemical descriptors of the amino acid index database (AAindex) [39,40] and sets of known natural GPCR ligand peptide retrieved from the bioactive peptide database (BioPD) [41]. We developed an automatic method to examine each descriptor in its ability to discriminate positive and negative sets (i.e. known GPCR peptidic ligands *vs* non-GPCR ligand peptides), and to find the optimal combination of the best discriminant descriptors for ligand/non-ligand peptide discrimination. Thus, PREDILIGAND use a model that selects GENE2PEP's novel natural peptides showing physico-chemical similarities with known GPCR natural ligand peptides (Figure 2).

SAGE Pattern Determination

A selection based on the putative localization and expression level of predicted peptides can help identify peptides with expression patterns similar to the ones of known GPCR peptidic ligands. An expression profile is a characterization of the relative quantity of every transcript that is produced in a cell type. One strategy that can be used to generate expression profiles is Serial Analysis of Gene Expression (SAGE). SAGE provides a functional profile of gene expression by determination of all the transcripts present within a cell. In addition, a differential expression pattern (healthy *vs* tumoral/infectious tissues) analysis will allow identification of key genes related to a specific disease or pathology [42]. The expression profiles obtained can be used either in a therapeutic perspective or in a process of identifying markers for diagnosis. Consequently, we have added a SAGE database to our bioinformatics platform. This database gathers more than 700 SAGE libraries representing 25 human tissues, 150 different human physio-pathological conditions and 20 different animal species. The location and expression level pattern of a predicted peptide is determined by extracting a tag from its EST and by determining the tag occurrence frequency in all of available SAGE libraries. Figure 3 shows the expression pattern of a putative gene encoding oGPCR peptidic ligands in different tissues and in a context of cancer. In this specific example, the putative gene is over-expressed in tumoral lung tissues. This suggests that the product of this gene may play a significant role in lung cancer. After SAGE pattern determination, the peptides are finally stored in G2Pdb, our relational database dedicated to novel natural GPCR-ligand-like peptides.

Predicted Peptide Storage and Consultation: G2Pdb

G2Pdb allows structuring and organizing the recording of the data generated by the processing of the transcriptome database flat files. G2Pdb content is browsable, thanks to a web user-friendly interface in which the user can submit selection criteria for getting a predicted peptide list. Table 1 summarizes the accessible criteria provided by G2Pdb. While some criteria come from annotation transfer from screened sequences (organism, taxon), others directly

Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

DOI: 10.1002/psc

arise from our bioinformatic prediction: ligand-like predicted status, PTMs, motif applied to predicted sequences and SAGE expression profiles. The example provided in Table 1 shows the criteria for the selection of a human peptide (size limitation from 7 to 15 amino acids) predicted as a GPCR-ligand-like peptide, with a *C*-terminal RFamide motif and for which the corresponding putative gene is five-fold more expressed in healthy brain tissues than in any other healthy tissue.

BIOINFORMATIC PLATFORM EXECUTION AND RESULTS

Our bioinformatics program pipeline has been run on both dbEST (release 150, October 2005) and UniProtKB/Swiss-Prot databases (release 51, October 2006). The execution of GENE2PEP and PREDILIGAND has led to the generation of ~10 000 peptide clusters regarding sequence similarity. In order to investigate the ability of our bioinformatics pipeline to predict expected known sequences, a list of peptides with ligand properties has been generated using the BioPD database [41] and we searched its members within our predicted peptides. This peptidic ligand list gathers 437 distinct sequences of which 326 are derived from the cleavage at basic residues of peptidic precursors (83%). GENE2PEP and PREDILIGAND allowed the prediction of 242 sequences of this subset (74%) and the conservation of 182 sequences out to the 326 expected ligand sequences (56%), respectively. In other words, our bioinformatic program pipeline has managed to recover more than half of currently well-known peptidic ligands. Table 2 shows a short list of known peptide precursor family for which the predictive programs have been able to find expected GPCR peptidic ligands

Table 1 Accessible criteria for putative oGPCR peptidic ligand selection

Criteria	Example
Organism	Homo sapiens
Taxon	Mammals
Organ	Brain
Tissue	Pituitary
Tissue differential overexpression	Five-fold
Physio-pathological state differential overexpression	Not assigned
Pathological state	Healthy
Ligand-like predicted state	True
Peptide length	6 < length < 16
PTM	*C-terminal amidation*
Motif	*C-terminal: RF*
Sequence reliability	3

Table 2 List of known natural peptide families (precursors and matured peptides) found after UniProtKB/Swiss-Prot flat-file processing

UniProtKB/Swiss-Prot peptide precursor family	Peptide example
Adrenomedullin	Adrenomedullin-2
Apelin	Apelin-28
Proopiomelanocortin (POMC)	Alpha-MSH (melanostimulating hormone)
Sauvagin/corticotropin-releasing factor/urotensin I	Corticotropin-releasing factor
Galanin	Galanin
Glucagon	Glucagon-like peptide 1
GnRH	GnRH-I
Bombesin/neuromedin B/ranatensin	Neuromedin B
Neuropeptide Y (NPY)	Neuropeptide Y
Neuropeptide B/W	Neuropeptide B-29
FMRFamide-related peptides	Neuropeptide FF
Orexin	Orexin-A
Opioid neuropeptides	Beta-neoendorphin
Somatostatin	Somatostatin-14
Tachykinin	Neurokinin A
Urotensin 2	Urotensin-2B

CONCLUSIONS AND PERSPECTIVES

Owing to their involvement in a large spectrum of physiological and pathological processes, GPCRs are preferred targets for the development of new therapeutics. Intensive research efforts for oGPCR deorphanization may lead to the discovery of new ligands, which may represent a source for new drugs. To achieve this, we have developed and designed an *in silico* bioinformatics screening relying on EST and protein databases. This bioinformatics approach has been implemented as a pipeline of predictive programs that simulates the biological events of peptide processing. In addition, this bioinformatics tool includes the selection of natural peptides sharing physico-chemical properties with known GPCR peptidic ligands. The peptide sequence files are also supplemented with specific data such as predicted PTMs and expression profiles. These data provide powerful selection criteria to determine the best peptide candidates for testing in functional assays. Peptides whose putative genes show differential physio-pathological expression patterns should be of particular interest. For example, Nash and Welch [43] observed that the KISS1 gene that encodes for metastin (kisspeptin-54), the ligand of the GPR54 receptor, is repressed or under-expressed in several different tumor types in metastatic diseases. As a potential metastatic gene suppressor, KISS1 is an important clinical target for the treatment and diagnosis of metastatic diseases [43]. This emphasizes the

Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

DOI: 10.1002/psc

importance of having in hand the differential expression of our predicted peptides to select the best candidates for their development as a drug or as a diagnosis marker. Using our bioinformatics-based discovery platform, novel GPCR ligand peptide candidates have been selected and produced chemically at our core facility in large quantities for activity screening. We are currently focusing on the establishment of lists gathering all predicted peptides from human genome that show expression patterns different from healthy and tumoral tissues. By having access to functional assays on GPCRs/oGPCRs, we will determine the potential activity of the retrieved new peptides in GPCR-mediated cancerous diseases. Consequently, our bioinformatics strategy will be adapted to the prediction of lytic peptides based on a physico-chemical model (PREDILYTIC) built with known anti-microbial peptides. This could lead to the identification of novel anti-microbial peptides which may represent good candidates for the development of new types of anti-infectious drugs.

Acknowledgements

We are very grateful to Dr Philippe Bulet for a critical reading of the manuscript.

We would like to acknowledge SkuldTech Company for providing acces to the SAGE technology. We would like to thank Isabelle Servant for reading the English version and Oseo ANVAR (financing of Karine Lessellier) and the French Research Ministry for financial support of Genepep.

REFERENCES

1. Drews J. Drug discovery: a historical perspective. *Science* 2001; **287**: 1960–1964.
2. Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, Bergmann JE, Gaitanaris GA. The G-protein-coupled receptor repertoires of human and mouse. *Proc. Natl. Acad. Sci. U.S.A.* 2003; **100**: 4903–4908.
3. Civelli O. GPCR deorphanizations: the novel, the known and the unexpected transmitters. *Trends Pharmacol. Sci.* 2005; **26**: 15–19.
4. Fargin A, Raymond JR, Lohse MJ, Kobilka BK, Caron MG, Lefkowitz RJ. The genomic clone G-21 which resembles a beta-adrenergic receptor sequence encodes the 5-HT1A receptor. *Nature* 1988; **335**: 358–360.
5. Bunzow JR, Van Tol HH, Grandy DK, Albert P, Salon J, Christie M, Machida CA, Neve KA, Civelli O. Cloning and expression of a rat D2 dopamine receptor cDNA. *Nature* 1988; **336**: 783–787.
6. Civelli O, Nothacker HP, Saito Y, Wang Z, Lin SH, Reinscheid RK. Novel neurotransmitters as natural ligands of orphan G-protein-coupled receptors. *Trends Neurosci.* 2001; **24**: 230–237.
7. Marchese A, George SR, O'Dowd BF. Cloning of G protein-coupled receptor genes. In *Identification and Expression of G Protein-Coupled Receptors*, Lynch KR (ed.). Wiley-Liss: 1998; 1–26.
8. Wise A, Jupe SC, Rees S. The identification of ligands at orphan G-protein coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* 2004; **44**: 43–66.
9. Civelli O, Saito Y, Wang Z, Nothacker HP, Reinscheid RK. Orphan GPCRs and their ligands. *Pharmacol. Ther.* 2006; **110**: 525–532.
10. Jiang Y, Luo L, Gustafson EL, Yadav D, Laverty M, Murgolo N, Vassileva G, Zeng M, Laz TM, Behan J, Qiu P, Wang L, Wang S, Bayne M, Greene J, Monsma F Jr, Zhang FL. Identification and characterization of a novel RF-amide peptide ligand for orphan G-protein-coupled receptor SP9155. *J. Biol. Chem.* 2003; **278**: 27652–27657.
11. Fredriksson R, Schiöth HB. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.* 2005; **67**: 1414–1425.
12. Anastasi A, Erspamer V, Bucci M. Isolation and structure of bombesin and alytesin, 2 analogous active peptides from the skin of the European amphibians Bombina and Alytes. *Experientia* 1971; **27**: 166–167.
13. Erspamer V, Erspamer GF, Inselvini M. Some pharmacological actions of alytesin and bombesin. *J. Pharm. Pharmacol.* 1970; **22**: 875–876.
14. McDonald TJ, Jornvall H, Nilsson G, Vagne M, Ghatei M, Bloom SR, Mutt V. Characterization of a gastrin releasing peptide from porcine non-antral gastric tissue. *Biochem. Biophys. Res. Commun.* 1979; **90**: 227–233.
15. Minamino N, Kangawa K, Matsuo H. Neuromedin B: a novel bombesin-like peptide identified in porcine spinal cord. *Biochem. Biophys. Res. Commun.* 1983; **114**: 541–548.
16. Spindel ER, Chin WW, Price J, Rees LH, Besser GM, Habener JF. Cloning and characterization of cDNAs encoding human gastrin-releasing peptide. *Proc. Natl. Acad. Sci. U.S.A.* 1984; **81**: 5699–5703.
17. Krane IM, Naylor SL, Helin-Davis D, Chin WW, Spindel ER. Molecular cloning of cDNAs encoding the human bombesin-like peptide neuromedin B. Chromosomal localization and comparison to cDNAs encoding its amphibian homolog ranatensin. *J. Biol. Chem.* 1988; **263**: 13317–13323.
18. Arlander SJ, Dong M, Ding XQ, Pinon DI, Miller LJ. Key differences in molecular complexes of the cholecystokinin receptor with structurally related peptide agonist, partial agonist, and antagonist. *Mol. Pharmacol.* 2004; **66**: 545–552.
19. Eipper BA, Mains RE. Peptide alpha-amidation. *Annu. Rev. Physiol.* 1988; **50**: 333–344.
20. Gentilucci L, Tolomelli A. Recent advances in the investigation of the bioactive conformation of peptides active at the micro-opioid receptor. Conformational analysis of endomorphins. *Curr. Top. Med. Chem.* 2004; **4**: 105–121.
21. In Y, Minoura K, Ohishi H, Minakata H, Kamigauchi M, Sugiura M, Ishida T. Conformational comparison of mu-selective endomorphin-2 with its C-terminal free acid in DMSO solution, by 1H NMR spectroscopy and molecular modeling calculation. *J. Pept. Res.* 2001; **58**: 399–412.
22. Walsh JH. Gastrin. In *Gut peptides*, Walsh JH, Dockray GJ (eds.). Raven Press: New York, 1994; 75–121.
23. Ouafik L, Sauze S, Boudouresque F, Chinot O, Delfino C, Fina F, Vuaroqueaux V, Dussert C, Palmari J, Dufour H, Grisoli F, Casellas P, Brunner N, Martin PM. Neutralization of adrenomedullin inhibits the growth of human glioblastoma cell lines in vitro and suppresses tumor xenograft growth in vivo. *Am. J. Pathol.* 2002; **160**: 1279–1292.
24. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis Of gene expression. *Science* 1995; **270**: 484–487.
25. Amblard M, Fehrentz JA, Martinez J, Subra G. Methods and protocols of modern solid phase peptide synthesis. *Mol. Biotechnol.* 2006; **33**: 239–254.
26. Thomsen W, Frazer J, Unett D. Functional assays for screening GPCR targets. *Curr. Opin. Biotechnol.* 2005; **16**: 655–665.
27. Boguski MS, Lowe TM, Tolstoshev CM. dbEST–database for "expressed sequence tags". *Nat. Genet.* 1993; **4**: 332–333.
28. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. The

Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

universal protein resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* 2006; **34**: (Database issue): D187–D191.
29. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* 2007; **35**: (Database issue): D21–D25.
30. Paetzel M, Karla A, Strynadka NC, Dalbey RE. Signal peptidases. *Chem. Rev.* 2002; **102**: 4549–4580.
31. Seidah NG, Chretien M. Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res.* 1999; **848**: 45–62.
32. Duncan EA, Brown MS, Goldstein JL, Sakai J. Cleavage site for sterol-regulated protease localized to a Leu-Ser bond in the lumenal loop of sterol regulatory element-binding protein-2. *J. Biol. Chem.* 1997; **272**: 12778–12785.
33. Seidah NG, Mowla SJ, Hamelin J, Mamarbachi AM, Benjannet S, Toure BB, Basak A, Munzer JS, Marcinkiewicz J, Zhong M, Barale JC, Lazure C, Murphy RA, Chretien M, Marcinkiewicz M. Mammalian subtilisin/kexin isozyme SKI-1: a widely expressed proprotein convertase with a unique cleavage specificity and cellular localization. *Proc. Natl. Acad. Sci. U.S.A.* 1999; **96**: 1321–1326.
34. Seidah NG, Benjannet S, Wickham L, Marcinkiewicz J, Jasmin SB, Stifani S, Basak A, Prat A, Chretien M. The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 2003; **100**: 928–933.
35. Mzhavia N, Pan H, Fricker LD, Devi LA. Characterization of Endothelin-converting enzyme-2, implication for a role in the nonclassical processing of regulatory peptides. *J. Biol. Chem.* 2003; **278**: 14704–14711.
36. Hook V, Yasothornsrikul S, Greenbaum D, Medzihradszky KF, Troutner K, Toneff T, Bundey R, Logrinova A, Reinheckel T, Peters C, Bogyo M. Cathepsin L and Arg/Lys aminopeptidase: a distinct prohormone processing pathway for the biosynthesis of peptide neurotransmitters and hormones. *Biol. Chem.* 2004; **385**: 473–480.
37. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 2000; **16**: 276–277.
38. Nielsen H, Krogh A. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB 6).* AAAI Press: Menlo Park, California, 1998; 122–130.
39. Kawashima S, Ogata H, Kanehisa M. AAindex: Amino acid index database. *Nucleic Acids Res.* 1999; **27**: 368–369.
40. Kawashima S, Kanehisa M. AAindex: amino acid index database. *Nucleic Acids Res.* 2000; **28**: 374.
41. Shi L, Zhang Q, Rui W, Lu M, Jing X, Shang T, Tang J. BioPD: a web-based information center for bioactive peptides. *Regul. Pept.* 2004; **120**: 1–3.
42. Maillard JC, Berthier D, Thevenon S, Piquemal D, Chantal I, Marti J. Efficiency and limits of the serial analysis of gene expression (SAGE) method: discussions based on first results in bovine trypanotolerance. *Vet. Immunol. Immunopathol.* 2005; **108**: 59–69.
43. Nash KT, Welch DR. The KISS1 metastasis suppressor: mechanistic insights and clinical utility. *Front. Biosci.* 2006; **11**: 647–659.

Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

DOI: 10.1002/psc