A EUROPEAN JOURNAL OF CHEMICAL BIOLOGY

CHEMBIOCHEM

SYNTHETIC BIOLOGY & BIO-NANOTECHNOLOGY

Accepted Article

Title: Novel concept for biosynthetic short neuropeptides: a rational theory based on experimental results for the missing pain-relief opioid endomorphin precursor gene

Authors: Ayami Matsushima, Jun Sese, and Kanako Koyanagi

To be cited as: *ChemBioChem* 10.1002/cbic.201900317

Link to VoR: http://dx.doi.org/10.1002/cbic.201900317

A Journal of

ChemPubSoc Europe

WILEY-VCH

www.chembiochem.org

ChemBioChem

Concept

Novel concept for biosynthetic short neuropeptides: a rational theory based on experimental results for the missing pain-relief opioid endomorphin precursor gene

Ayami Matsushima[1]*, Sese Jun[2,3,4], Kanako Koyanagi[5]

[1]Laboratory of Structure-Function Biochemistry, Department of Chemistry, Faculty and Graduate School of Science, Kyushu University, Fukuoka 819-0395, Japan, [2]Artificial Intelligence Research Center, AIST, 2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan, [3]AIST-Tokyo Tech RWBC-OIL, 2-12-1 Okayama, Meguro-ku, Tokyo 152-8550, Japan, [4]Humanome Lab, Inc. 2-4-10 Tsukiji, Chuo-ku, 104-0045, Tokyo, Japan, and [5]Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

*Author to whom correspondence should be addressed; E-Mail: ayami@chem.kyushu-univ.jp; Tel: +81-92-802-4159/Fax: +81-92-802-4126.

ABSTRACT

Endomorphins are neuropeptides which bind strongly to μ-opioid receptors and are considered to play important roles in pain modulation and other biological functions. Two endomorphins have been identified to date, endomorphine-1 and -2; both are tetrapeptides and differ by only a single amino acid difference at their third position. Both peptides were isolated from bovine brains; however, their precursor genes have not been identified. In this study, we found a nucleotide sequence corresponding to the endomorphin-1 peptide in an expressed sequence tag (EST) database and cloned a preproendomorphin-like precursor peptide from human brain cDNAs. The cDNA consists of nucleotide sequences of two already annotated predicted genes, and the putative peptide differs by one amino acid difference from the isolated endomorphin peptides. We propose here that the possibility that unknown short proteins or peptide precursors may be missed by automated gene prediction programs based on similarities of known protein sequences. We describe a novel concept of how to be produced endomorphins from a similar peptide. The oxidatively modified base might provide a clue for understanding discrepancies between nucleotide sequences on the genome and those on cDNAs.

***Keywords*:** 8-oxoguanine, endomorphin, opioid, peptide precursor

Introduction

Narcotic analgesics, also known as opioids, accounted for 36% of the major pharmaceutical markets of pain modulation[1]. Opioid peptides are intrinsic neuropeptides enzymatically processed from their precursor proteins. Three precursor proteins of intrinsic opioid peptides have been reported: preproopiomelanocortin (prePOMC), preproenkephalin, and preprodynorphin[2,3]. These three precursors include β- and γ-lipotropins, β-endorphin, adenocorticotropic hormone (ACTH), and α-, β-, and γ-melanocyte stimulating hormones (MSHs) in prePOMC; Met-enkephalin, Leu-enkephalin, and their related peptides in preproenkephalin; and α- and β-neoendorphins, dynorphin A, dynorphin B, and related peptides in preprodynorphin[3]. Endomorphin-1 (Tyr-Pro-Trp-Phe-NH_2) and endomorphin-2 (Tyr-Pro-Phe-Phe-NH_2) have been isolated from bovine[4] and human brains[5] and exhibit high binding ability for the μ-opioid receptor, to which morphine binds strongly; however, their precursor genes have not been cloned and the exact amino-acid sequences of the precursor proteins have not been identified. Therefore, the biosynthesis of endomorphins attracts many researchers. A de novo synthetic route of endomorphins involving reversal of peptidehydrolase catalytic activities are proposed based on the experiments by injection of [^{3}H]-Tyr-Pro into rat brains[6]. It was suggested that enzymatic conjugation of smaller peptide fragments by a comparable mechanism to the production of prokaryotic non-ribosomal peptide; however, such a mechanism is not reported in ineukaryotic cells, and thus the proposed endomorphin de novo biosynthetic mechanism still needs more robust experimental confirmations. Endomorphins are now thought to be good candidates for pain modulations because of their high biological activity, and improvements of their metabolic stability and/or their ability to cross the blood-brain barrier are needed and well studied[7–10]. All three of the previously reported precursor protein genes of opioid peptides encode multiple neuropeptides; therefore, the preproendomorphin gene would be expected to include novel bioactive peptides with potential utility for pain control. In

the preceding years, we and many researchers have made efforts to find the preproendomorphin genes. In the course of this effort, no one has reported the precursor gene encoding the amino acid sequences of endomorphin-1 and -2, although several groups, including ours, have identified nucleotide sequences that encodes similar peptide sequences with endomorphins[11]. The aim of this paper is to advocate a novel concept for identifying the missing preproendomorphin gene and putative peptide precursors. We suggest here the possibility of transcribing short peptides from putative protein genes already annotated by gene prediction programs with some epigenetic modifications.

Endomorphins were first discovered in bovine brains

Endomorphin-1 and endomorphin-2 are the experimentally isolated opioid peptides that show the highest affinity and selectivity for the μ-opioid receptor among mammalian endogenous opioid peptides. Both endomorphin-1 and endomorphin-2 were first discovered in bovine brain and then later in human brain. In the case of bovine brains, it was reported that ~200 ng of peptide including 75–80 % of endomorphin-1 and 15–20 % of endomorphin-2, corresponding to 2.1 pmol/g, were isolated from of 157 g of bovine frontal cortex[5]. From the human brains, 70-fold higher contents of these peptides were purified, and it was reported that the differences in concentration between the bovine and human peptides were attributed to a variety of factors[5]. After the isolation of the endomorphin peptides, a variety of biochemical, immunohistochemical, pharmacological, and behavioral researches were performed using many different approaches[12–14]. Endomorphin-1 and endomorphin-2 activate both $G_{\alpha i1}$- and $G_{\alpha oA}$-type G_{α} proteins via the μ-opioid receptor [15]. The distributions of endomorphin-1 and endomorphin-2 have been analyzed in detail using anti-endomorphins antibodies[3,13]. Both peptides produce transient antihyperalgesic and analgesic effects associated with an unresolved tolerance mechanism. Animal experiments using rodent models have suggested that endomorphins may reduce anxiety and depression[16,17]. Endomorphin

analogs that reduce respiratory depression, tolerance, abuse liability, and motor impairment have veen developed for use as new analgesic drugs[9,18]. Many researches using endomorphin peptides have indicated that these peptides may be endogenous ligands for μ-opioid receptors; however, their precursors genes have yet to be identified[19].

Our database search identified a peptide highly similar to endomorphins

We have performed a database search to find the putative endomorphin precursor peptides. The target nucleotide sequences were designed based on the following two points. (1) C-terminal glycines in the precursors of amide peptides are essential as sources for nitrogens of amide groups, because C-terminal amidation in natural peptide biosynthesis in mammals is mediated by peptidylglycine α-amidating monooxygenase [20–22]. (2) As seen in the three known opioid prepropeptide, neuropeptides are liberated by enzymes acting on their basic amino acids. The finally designed target sequences are all nucleotide sequences corresponding to the peptide sequences of Lys-Tyr-Pro-Trp-Phe-Gly-Lys, Lys-Tyr-Pro-Phe-Phe-Gly-Lys, Lys-Tyr-Pro-Trp-Phe-Gly-Arg, Lys-Tyr-Pro-Phe-Phe-Gly-Arg, Arg-Tyr-Pro-Trp-Phe-Gly-Lys, Arg-Tyr-Pro-Phe-Phe-Gly-Lys, Arg-Tyr-Pro-Trp-Phe-Gly-Arg, and Arg-Tyr-Pro-Phe-Phe-Gly-Arg. We searched the mammalian expressed sequence tag (EST) database for all of nucleotide sequences corresponding to these peptides. The finally identified EST had accession No. AI352151.1 and was defined as qr08h11.x1 soares total fetus Nb2HF8 9w Homo sapiens cDNA clone IMAGE:1940325 3', mRNA sequence. This sequence contained the nucleotides corresponding to the amino acid sequence of Arg-Tyr-Pro-Trp-Phe-Gly-Arg. The nucleotide sequence of accession No. AI352151.1 partially matched the genome sequences of Homo sapiens chromosome 8; however, the amino acid sequence encoded by this nucleotide region is Arg-Tyr-Pro-Gly-Phe-Gly-Arg.

cDNA cloning of a putative preproendomorphin-1 gene and binding ability of an endomorphine-1-like peptide

To confirm that the nucleotide sequence determined as described above was the actual nucleotide sequence for this region, we performed cDNA cloning using human brain cDNA, and determined the expressed mRNA sequence in the human brains. The putative amino acid sequence of the encoded peptide was Arg-Tyr-Pro-Gly-Phe-Gly-Arg; however, against our expectation, the identified nucleotide sequence was partly different from EST accession No. AI352151.1. Alignment of the nucleotide sequences of our identified full-length putative preproendomorphin-1 and the EST accession No. AI352151.1 showed that these are transcriptional variants, because the nucleotide sequence of EST lacks a 159 bp region in its middle part compared to that of the putative preproendomorphin-1 (Figure 1A and 1B). The identified nucleotide sequence encodes 94 amino acids from the first methionine (Met) after the stop codon~~;~~. The identified sequence has an adenosine (A) at the -3 position from the first Met, which fulfills a Kozak rule and this is important for gene translation[23,24]. We used the SignalP program to predict whether the putative preproendomorphin-1 has a signal peptide[25,26]. The precursor contains two other Mets in addition to the first one, and thus this nucleotide sequence has three possibilities for the initiating Met. The results indicated that the N-terminal 21 amino acids from the first Met are predicted to be a signal peptide that directs the protein across the endoplasmic reticulum membrane in eukaryotes[27,28] (Figure 2). The putative amino acid sequence encoded by this precursor would be Tyr-Pro-Gly-Phe-NH_2 (Figure 3). Many structure-function activity studies of endomorphins have been reported; however, there has been no report describing the binding ability of this peptide sequence to μ opioid receptors. In this study, we chemically synthesized Tyr-Pro-Gly-Phe-NH_2, and confirmed that this peptide showed extremely weak binding ability (>10,000 nM) to μ opioid receptors and no binding ability to δ and κ opioid receptors (data not shown).

Figure 1

Figure 2

Figure 3

The putative preproendomorphin-1-like gene consists of the nucleotide sequences of two adjacent predicted proteins

Using the obtained nucleotide sequence as a query, we searched the nucleotide collection database using the nucleotide BLAST program[29,30], and the obtained nucleotide sequence was a combination of the nucleotide sequences of two adjacent predicted proteins, the predicted Homo sapiens DDHD domain containing 2 and predicted Pan troglodytes nuclear receptor binding SET domain protein 3. The sequence was coincident with a combination of the 3 annotated genes: (1) PREDICTED Homo sapiens DDHD domain-containing 2, transcript variant X4, misc_RNA, (2) Homo sapiens DDHD domain containing 2, transcript variant 9, non-coding RNA, and (3) PREDICTED Pan troglodytes nuclear receptor binding SET domain protein 3, transcript variant X6, mRNA (Figure 4A and 4B). Because not only the nucleotide sequence found in the EST database but also that of our obtained in human brain cDNAs is composed of these two adjacent predicted genes, this putative preproendomorphin-1-like gene is transcribed in human brains as a mRNA.

Figure 4

Oxidization of a guanine allows the delivery of peptides having the same amino acid sequences as the isolated endomorphin

Nucleotides in the genome are naturally modified and/or damaged by genotoxic stresses such as those caused by oxidative stress-related molecules, and cells have repair systems to counteract the adverse effects of DNA damage in some cases causing cancers[31–33]. The oxidatively damaged base 8-oxoguanine is caused by cellular oxidative stress due to the oxidation guanine heterocycle, and this mutation leads to G→T transversion[34]

(Figure 5A). The nucleotide sequence of cloned endomorphin-1-like peptide is TAC (corresponding to a codon of Tyr)-CCT(Pro)- GGG(Gly)-TTT(Phe). When the first nucleotide of the codon of Gly is T, the resulting peptide is TAC(Tyr)-CCT(Pro)-**T**GG(**Trp**)-TTT(Phe) (Figure 5B), precisely the same amino acid sequence of the isolated endomorphin-1. If a special condition could be established such that the guanine of endomorphin-1-like gene would be oxidized and not repaired, the neuropeptide possessing the same amino acid sequence as the isolated endomorphine-1 should be delivered naturally. Even this region might contain positions highly sensitive to oxidation, and the peptide TAC(Tyr)-CCT(Pro)-**TTT**(**Phe**)-TTT(Phe) corresponding to endomorphin-2 might be produced from the identical precursor gene. Typical analysis algorithms of next generation sequencing (NGS) technologies use consensus nucleotide sequences to deal with sequencing errors, and thus might overlook the minor forms of expressed mRNA. Neuropeptides generally bind strongly to their receptors, and therefore only a small number of peptides is usually sufficient to display their bioactivity. Endomorphins have strong binding ability to μ-opioid receptors, more potent than that of exogenous ligand morphine[13,19]. If endomorphins are delivered only under certain lethal conditions, and are not produced normally, this might explain why, historically, endomorphins have been isolated from a large amount of bovine brains, although no genes encoding the same amino acid sequences have yet been identified.

EXPERIMENTAL SECTION

In silico cloning

All of the 21 mer nucleotide sequences corresponding to the amino acid sequences of endomorphin-1 and endomorphin-2 with basic amino acids (Arg or Lys) in the N-terminus and amidation sites (Gly-Arg or Gly-Lys) in the C-terminus were searched as queries from an EST database using the blastn program under a word size setting of 8.

cDNA cloning

Human brain, whole, QUICK-Clone cDNA was purchased from TAKARA Bio Inc. (Shiga, Japan). Firstly, PCR was performed using gene-specific primers designed based on ESR accession no. AI352151.1 (5'-TGCAGGCAGAAGCAGTTGGT-3' and 5'-TGAAGATACAGTATTGCTCGT-3'). Furthermore, nested PCR was conducted using gene-specific primers (5'-TTTTATTGTGCTTCAGTGCCA-3' and 5'-TGAAGATACAGTATTGCTCGT-3'). The obtained PCR products for an endomorphin-like peptide precursor were subcloned into T-vector pMD20 (TAKARA) and nucleotide sequences were analyzed (GenBank accession No. LC481596).

Peptide synthesis and receptor-binding assay

The endomorphin-1-like peptide H-Tyr-Pro-Gly-Phe-NH_2 was synthesized by the manual solid phase peptide synthesis method by the Fmoc-chemistry strategy as previously reported[35]. The binding ability of this peptide was analyzed by a competitive binding assay using [H^3]DAMGO against μ-opioid receptors provided as membrane preparation obtained from COS-7 cells that transiently expressed μ-opioid receptors[36]. MALDI-TOF *m*/*z* calcd for $C_{25}H_{32}N_5O_5+H^+$ (ave): 482.56; found: 482.74.

FIGURE LEGENDS

Figure 1. Comparison between the nucleotide sequences of the cloned putative endomorphin-1-like precursor and those of EST accession No. AI352151.1. A) The results of the alignment of these two sequences by the NCBI blastn program (https://blast.ncbi.nlm.nih.gov/Blast.cgi) indicated that the cloned cDNA contains the same nucleotide sequences with the insertion of 159 bps in its middle region. **B)** Dot matrix view of the results of the pairwise alignment showed that the nucleotide sequences of the cloned cDNA and EST accession No. AI352151.1 were almost identical, and EST accession No. AI352151.1 was divided by the inserted sequence of the cloned cDNAs.

Figure 2. Prediction of a signal peptide of the putative endomorphin-1-like precursor peptide. The first 21 amino acids from the initial Met are predicted to be a secretion signal peptide from Eukarya with a high likelihood score of 0.9197. The probability scores of a cleavage site between amino acid positions 22/23 and 24/25 are 0.4909 and 0.3616, respectively. The probability scores at each amino acid position are indicated as a secretion signal peptide (red line), cleavage site (green line), and other (yellow line).

Figure 3. The nucleotide sequence of the cloned putative endomorphin-1-like precursor and its translated amino acid sequence. The cloned cDNA encodes a 94 amino acid precursor peptide. This precursor has two putative amidation sites Gly-Arg. The amino acid sequence of the encoded peptide appears to be Tyr-Pro-Gly-Phe-NH_2, in which the third amino acid residue is different from that in endomorphin-1: Tyr-Pro-Trp-Phe-NH_2.

Figure 4. The nucleotide sequence of the cloned putative endomorphin-1-like precursor consists of two already annotated adjacent genes predicted to be proteins. **A)** The nucleotide sequence consists of predicted Pan troglodytes nuclear receptor binding SET domain protein 3 (NSD3) and predicted Homo sapiens DDHD domain-containing 2 (DDHD) located on human chromosome 8. Green bars indicate the gene locations and purple bars show the transcribed regions in each variant. Gray lines indicate possible introns. **B)** Structures of the nucleotide sequences of (a) the cloned putative endomorphin-1-like precursor, (b) EST accession No. AI352151.1, and (c) the results obtained from a NCBI BLAST search by the blastn program.

10.1002/cbic.201900317

263 **Figure 5. Hypothesized pathway for delivery of the endomorphin-1 peptide from the**
264 **identified cDNA.** Because 8-Oxoguanines cause G→T transversions, under certain
265 specific conditions this mutation might produce the amino acid sequence of the isolated
266 endomorphin-1 as a functional neuropeptide from the cloned endomorphin-1-like
267 precursor gene.
268
269

Acknowledgements

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI grant numbers 26670288, JP17H01881 and JP18K19147 (to A.M.).

Competing interests

The authors declare no competing interests.

Keywords

8-oxoguanine, endomorphin, opioid, peptide precursor

References

[1] I. Melnikova, *Nat. Rev. Drug Discov.* **2010**, *9*, 589-590.

[2] D. T. Krieger, A. Liotta, M. J. Brownstein, *Brain Res.* **1977**, *128*, 575-579.

[3] H. Akil, S. J. Watson, E. Young, *Annu. Rev. Neurosci.* **1984**, *7*, 223-255.

[4] J. E. Zadina, L. Hackler, L. J. Ge, A. J. Kastin, *Nature* **1997**, *386*, 499-502.

[5] L. Hackler, J. E. Zadina, L. J. Ge, A. J. Kastin, *Peptides* **1997**, *18*, 1635-1639.

[6] A. Z. Rónai, E. Szemenyei, E. Kató, L. Kocsis, G. Orosz, M. Al-Khrasani, G. Tóth, *Regul. Pept.* **2006**, *134*, 54-60.

[7] E. R. Siuda, R. Carr, D. H. Rominger, J. D. Violin, *Curr. Opin. Pharmacol.* **2017**, *32*, 77-84.

[8] T. Günther, P. Dasgupta, A. Mann, E. Miess, A. Kliewer, S. Fritzwanker, R. Steinborn, S. Schulz, *Br. J. Pharmacol.* **2018**, *175*, 2857-2868.

[9] J. E. Zadina, M. R. Nilges, J. Morgenweck, X. Zhang, L. Hackler, M. B. Fasold, *Neuropharmacology* **2016**, *105*, 215-227.

[10] R. De Marco, A. Bedini, S. Spampinato, L. Comellini, J. Zhao, R. Artali, L. Gentilucci, *J. Med. Chem.* **2018**, *61*, 5751-5757.

[11] M. E. Matus-Ortega, P. Leff Gelman, J. C. Calva-Nieves, A. Flores-Zamora, A. Salazar-Juárez, C.-A. Torner-Aguilar, G. Gamba, P. De Los Heros, B. Peng, J. E. Pintar, H. S. Gompf, C. N. Allen, B. Antón-Palma, *FEBS Lett.* **2017**, *591*, 1627-1636.

[12] G. A. Olson, R. D. Olson, A. L. Vaccarino, A. J. Kastin, *Peptides* **1998**, *19*, 1791-1843.

[13] Z.-H. Gu, B. Wang, Z.-Z. Kou, Y. Bai, T. Chen, Y.-L. Dong, H. Li, Y.-Q. Li, *Neurosignals* **2018**, *25*, 98-116.

[14] R. J. Bodnar, *Peptides* **2018**, *101*, 167-212.

[15] Z. Saidak, K. Blake-Palmer, D. L. Hay, J. K. Northup, M. Glass, *Br. J. Pharmacol.* **2006**, *147*, 671-680.

[16] S. Rotzinger, D. A. Lovejoy, L. A. Tan, *Peptides* **2010**, *31*, 736-756.

[17] J. E. Zadina, M. R. Nilges, J. Morgenweck, X. Zhang, L. Hackler, M. B. Fasold, *Neuropharmacology* **2016**, *105*, 215-227.

[18] A. K. Feehan, J. Morgenweck, X. Zhang, A. T. Amgott-Kwan, J. E. Zadina, *J. Pain* **2017**, *18*, 1526-1541.

[19] J. Fichna, A. Janecka, J. Costentin, J.-C. Do Rego, *Pharmacol. Rev.* **2007**, *59*, 88-123.

[20] B. A. Eipper, L. P. Park, I. M. Dickerson, H. T. Keutmann, E. A. Thiele, H. Rodriguez, P. R. Schofield, R. E. Mains, *Mol. Endocrinol.* **1987**, *1*, 777-790.

[21] I. Kato, H. Yonekura, M. Tajima, M. Yanagi, H. Yamamoto, H. Okamoto, *Biochem. Biophys. Res. Commun.* **1990**, *172*, 197-203.

[22] S. N. Perkins, E. J. Husten, B. A. Eipper, *Biochem. Biophys. Res. Commun.* **1990**, *171*, 926-932.

[23] M. Kozak, *J.Cell Biol.* **1989**, *108*, 229-241.

[24] M. Kozak, *Gene* **2002**, *299*, 1-34.

[25] H. Nielsen, J. Engelbrecht, S. Brunak, G. Von Heijne, *Protein Eng.* **1997**, *10*, 1-6.

[26] J. J. Almagro Armenteros, K. D. Tsirigos, C. K. Sønderby, T. N. Petersen, O. Winther, S. Brunak, G. Von Heijne, H. Nielsen, *Nat. Biotech.* **2019**, *37*, 420-423.

[27] G. Von Heijne, *J. Mol. Biol.* **1985**, *184*, 99-105.

[28] G. Von Heijne, *J. Mol. Biol.***1990**, *115*, 195-201.

[29] Z. Zhang, S. Schwartz, L. Wagner, W. Miller, *J. Comput. Biol.* **2000**, *7*, 203-214.

[30] A. Morgulis, G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala, A. A. Schäffer, in *Bioinformatics* **2008**, *24*, 1757-1764.

[31] M. Christmann, B. Kaina, *Nucleic Acids Res.* **2013**, *41*, 8403-8420.

[32] N. C. Bauer, A. H. Corbett, P. W. Doetsch, *Nucleic Acids Research* **2015**, *43*, 10083-10101.

[33] S. A. Roberts, D. A. Gordenin, *Nat. Rev. Cancer* **2014**, *14*, 786-800.

[34] Y. Ding, A. M. Fleming, C. J. Burrows, *J. Am. Chem. Soc.* **2017**, *139*, 2569-2572.

[35] A. Matsushima, H. Nishimura, Y. Matsuyama, X. Liu, T. Costa, Y. Shimohigashi, *Biopolymers* **2016**, *106*, 460-469.

[36] T. Honda, N. Shirasu, K. Isozaki, M. Kawano, D. Shigehiro, Y. Chuman, T. Fujita, T. Nose, Y. Shimohigashi, *Bioorg. Med. Chem.* **2007**, *15*, 3883-3888.

10.1002/cbic.201900317

FIGURES

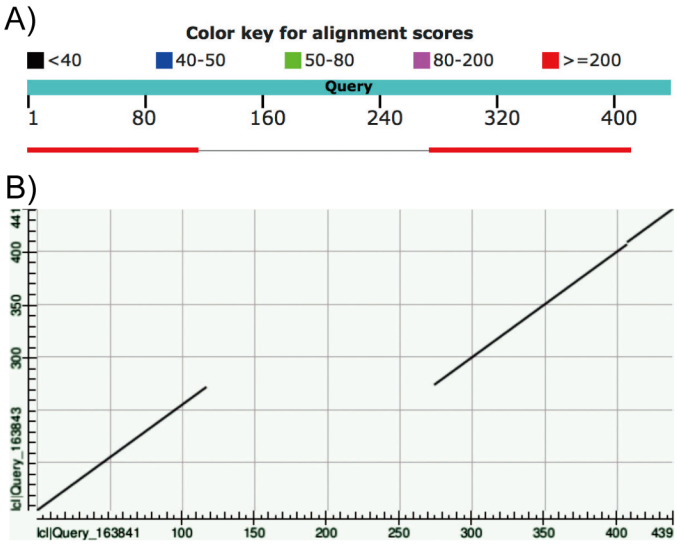

Figure 1

Figure 1. Comparison between the nucleotide sequences of the cloned putative endomorphin-1-like precursor and those of EST accession No. AI352151.1. A) The results of the alignment of these two sequences by the NCBI blastn program (https://blast.ncbi.nlm.nih.gov/Blast.cgi) indicated that the cloned cDNA contains the same nucleotide sequences with the insertion of 159 bps in its middle region. **B)** Dot matrix view of the results of the pairwise alignment showed that the nucleotide sequences of the cloned cDNA and EST accession No. AI352151.1 were almost identical, and EST accession No. AI352151.1 was divided by the inserted sequence of the cloned cDNAs.

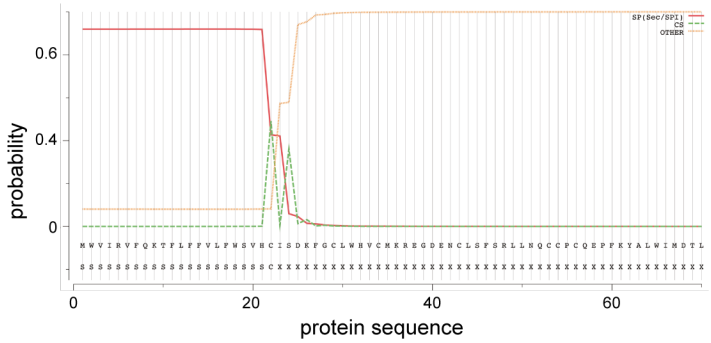

Figure 2

Figure 2. Prediction of a signal peptide of the putative endomorphin-1-like precursor peptide. The first 21 amino acids from the initial Met are predicted to be a secretion signal peptide from Eukarya with a high likelihood score of 0.9197. The probability scores of a cleavage site between amino acid positions 22/23 and 24/25 are 0.4909 and 0.3616, respectively. The probability scores at each amino acid position are indicated as a secretion signal peptide (red line), cleavage site (green line), and other (yellow line).

```
120 TAC ACT CCC AAA ATG TGG GTA ATC CGT GTC TTT CAA AAA ACA TTT CTG TTT TTT GTT TTG
  1                 M   W   V   I   R   V   F   Q   K   T   F   L   F   F   V   L
180 TTT TGG TCA GTC CAT TGC ATA AGT GAC AAG TTT GGG TGC TTG TGG CAC GTA TGT ATG AAG
 17  F   W   S   V   H   C   I   S   D   K   F   G   C   L   W   H   V   C   M   K
240 CGG GAG GGG GAT GAG AAT TGC CTG TCC TTC AGT AGG CTG CTC AAC CAA TGC TGC CCG TGC
 27  R   E   G   D   E   N   C   L   S   F   S   R   L   L   N   Q   C   C   P   C
300 CAA GAG CCT TTT AAA TAT GCT TTG TGG ATT ATG GAT ACA TTA AGC AGC CAT AGA TGG GTC
 47  Q   E   P   F   K   Y   A   L   W   I   M   D   T   L   S   S   H   R   W   V
360 ATT TTT ACT GTA AAG GCT GAT CAA GGA AGA TAC CCT GGG TTT GGT AGA TCT CTT TGA GGA
 67  I   F   T   V   K   A   D   Q   G   R   Y   P   G   F   G   R   S   L   *
```

Figure 3

Figure 3. The nucleotide sequence of the cloned putative endomorphin-1-like precursor and its translated amino acid sequence. The cloned cDNA encodes a 94 amino acid precursor peptide. This precursor has two putative amidation sites Gly-Arg. The amino acid sequence of the encoded peptide appears to be Tyr-Pro-Gly-Phe-NH_2, in which the third amino acid residue is different from that in endomorphin-1: Tyr-Pro-Trp-Phe-NH_2.

10.1002/cbic.201900317

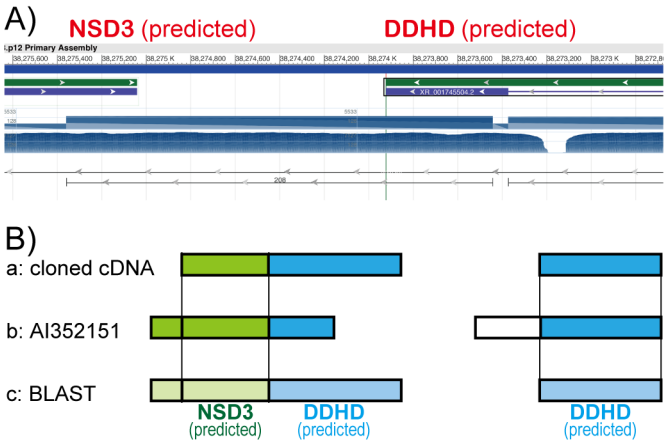

Figure 4

Figure 4. The nucleotide sequence of the cloned putative endomorphin-1-like precursor consists of two already annotated adjacent genes predicted to be proteins. **A)** The nucleotide sequence consists of predicted Pan troglodytes nuclear receptor binding SET domain protein 3 (NSD3) and predicted Homo sapiens DDHD domain-containing 2 (DDHD) located on human chromosome 8. Green bars indicate the gene locations and purple bars show the transcribed regions in each variant. Gray lines indicate possible introns. **B)** Structures of the nucleotide sequences of (a) the cloned putative endomorphin-1-like precursor, (b) EST accession No. AI352151.1, and (c) the results obtained from a NCBI BLAST search by the blastn program.

10.1002/cbic.201900317

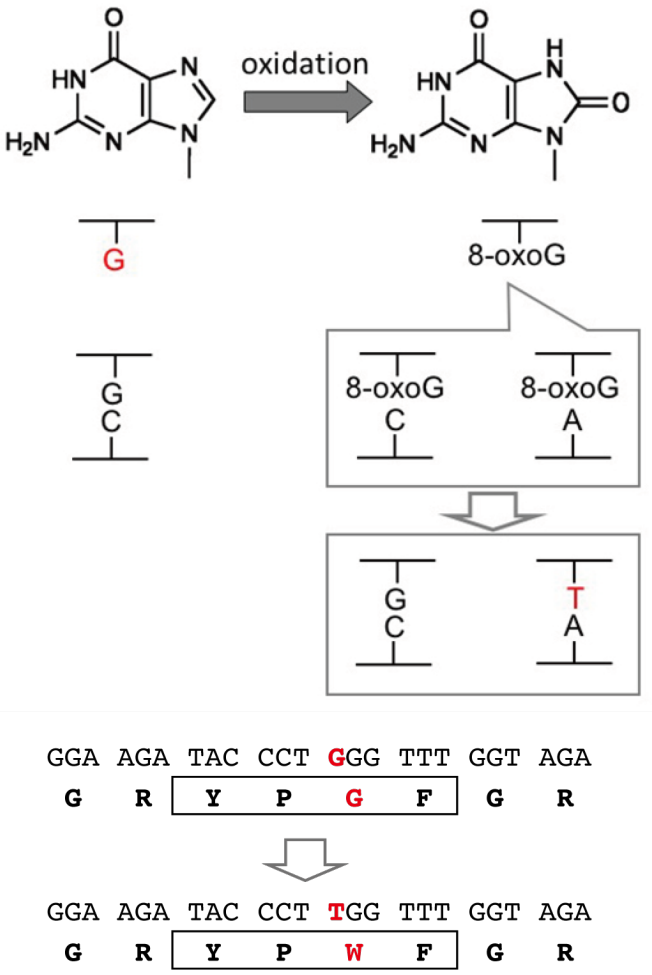

Figure 5

Figure 5. Hypothesized pathway for delivery of the endomorphin-1 peptide from the identified cDNA. Because 8-Oxoguanines cause G→T transversions, under certain specific conditions this mutation might produce the amino acid sequence of the isolated endomorphin-1 as a functional neuropeptide from the cloned endomorphin-1-like precursor gene.