Research Article

NEUROPEPTIDE PREDICTED IN EMBERIZA BRUNICEPS USING COMPUTATIONAL TOOLS

Singh S*, Gautam B, Jain P. A, Yadav P.K and Farmer R.
Department of Computational Biology & Bioinformatics, JSBBE, SHIATS, Allahabad-211007, India.
*Corresponding author's E-mail: satendralike@gmail.com

ABSTRACT

The present study was carried out to predict the function of recently cloned nucleotide sequence of *Emberiza bruniceps*. The amino acid sequence of cloned nucleotide was retrieved from the GenBank. After using BLASTP search, conserve domain identification tool and motif search for the selected amino acid sequence it was found that the sequence has significant identify with many sequences in the biological databases. Further the sequence has also shown the presence of conserve domain in the protein of some organisms. All these finding leads to identification of amino acid sequence as the neuropeptide which is a regulator of pancreatic and gastrointestinal functions. This prediction was further supported by identification of a pancreatic hormone motif in the amino acid sequence. The identified motif was also present in many of available protein sequences and structures in the PDB.

Keywords: *Emberiza bruniceps*, Neuropeptide, Conserve domain, Motif.

INTRODUCTION

The *Emberiza bruniceps* is a bird in the bunting family Emberizidae. It is migratory, wintering in India. Red-headed Bunting breeds in open scrubby areas including agricultural land. Its natural food consists of insects when feeding young and otherwise seeds. The migratory pattern of this bird is affected by different factors which include climate conditions and availability of food[1, 2]. Recently a nucleotide sequence was cloned under the starving condition. The nucleotide sequence was submitted to GenBank, while the function of the sequences remained unknown. The present study has predicted the possible function of the translated nucleotide sequence with the help of various computational tools.

MATERIALS AND METHODS

The translated nucleotide in the form of amino acid sequence ADD83204.1 was retrieved from the GenBank in the FASTA format. BLASTP search was carried out to find identical or similar sequence in the non-redundant protein sequence database. The search was performed at E-Value 0.01 and the database used for conserve domain is CDS search with tool cddv2.25. The retrieved amino acid sequence was further used to identify the presence of conserve domain with the help of the conserve domain identification tool present on NCBI. The same sequence was again used to predict possible motif in the sequence by using motif search tool against the PROSITE pattern database[3, 4].

RESULTS AND DISCUSSION

The BLASTP search for the query sequence has identified several identical and similar protein sequences occurring in wide range of organisms. Some of those significant proteins are listed with their accession number in Table 1.

Almost all such proteins in different organisms occur as neuropeptide.

Table 1: Similar proteins for query sequence ADD83204.1

S.No	Accession Number identified by BLASTP search	Type of Protein
1	XP_002193048.1	Neuropeptide
2	NP_990804.1	Neuropeptide
3	ACB86930.1	Neuropeptide
4	XP_002818184.1	Neuropeptide
5	NP_000896.1	Neuropeptide
6	XP_001159002.1	Neuropeptide
7	NP_001027986.1	Neuropeptide
8	XP_532492.2	Neuropeptide
9	NP_001153758.1	Neuropeptide
10	AAT66407.1	Neuropeptide
11	NP_001072530.1	Neuropeptide
12	AAA59945.1	Neuropeptide
13	NP_075945.1	Neuropeptide
14	Q9PW68.1	Neuropeptide
15	NP_001014845.1	Neuropeptide
16	AAH80407.1	Neuropeptide
17	NP_001161232.1	Neuropeptide
18	AAC69886.1	Neuropeptide
19	Q9PTA0.1	Neuropeptide
20	NP_001116379.1	Neuropeptide
21	XP_002748087.1	Neuropeptide
22	XP_001113958.1	Neuropeptide
23	ABW05041.1	Neuropeptide

Since sequence similarity does not give accurate picture of the function of any protein, the conserve domain in the query sequence was also identified. The same conserve domain was present in 213 proteins other proteins. Some

of these conserve domain were presented in Figure 1. The conserve domain was identified as pancreatic hormone domain in the query sequence. Pancreatic hormone domain is a regulator of pancreatic and gastrointestinal functions; neuropeptide Y (NPY)b, peptide YY (PYY), and pancreatic polypeptide (PP) are closely related. The two significant regions identified in query sequence are receptor binding site and dimerization interface region. The receptor binding site is further confirmed by the presence of arginines in the query sequence which is required for receptor binding[5].
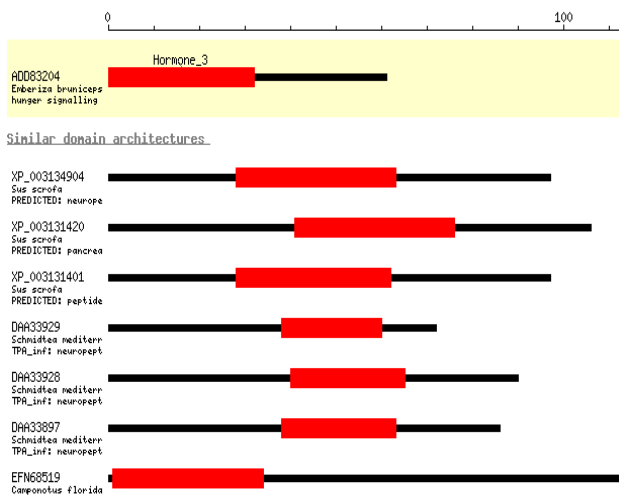

Figure 1: Representation of conserve domains in different sequences.
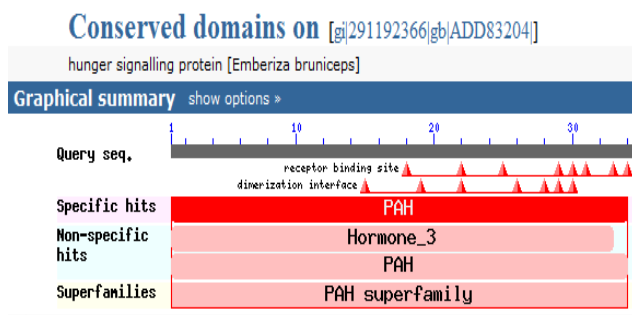

Figure 2: Significant Conserve domain sites

Receptor binding site on conserved domain PAH. 8 of 8 of the residues and dimerization interface on conserved domain PAH. 7 of 7 of the residues that compose this conserved feature have been mapped to the query sequence as presented in Figure 2[6].

The presence of pancreatic hormone domain in the query sequence was further verified by the identification of motif in the query sequence. The residues of amino acid in the motif and the regular expression are present in Figure 3. The motif identified in query sequence was again identified as pancreatic hormone motif[7,8], which confirms the previous function prediction of the query sequence.

Figure 3: Identified motif in the sequence and its regular expression

The identified motif was also found to be present in 13 proteins in the PDB. These proteins with their PDB id are presented in Table 2.

Table 2: List of structurally known proteins having Pancreatic Hormone Motif

S.No.	PDB ID	Proteins having Pancreatic Hormone Motif
1	1BBA	SEQUENCE-SPECIFIC 1H NMR ASSIGNMENTS AND SOLUTION STRUCTURE OF BOVINE PANCREATIC POLYPEPTIDE
2	1F8P	PORCINE NEUROPEPTIDE Y BOUND TO DPC MICELLES
3	1K8V	THE NMR-DERIVED CONFORMATION OF NEUROPEPTIDE F FROM MONIEZIA EXPANSA
4	1LJV	BOVINE PANCREATIC POLYPEPTIDE BOUND TO DPC MICELLES
5	1PPT	X-RAY ANALYSIS (1.4-ANGSTROMS RESOLUTION) OF AVIAN PANCREATIC POLYPEPTIDE. SMALL GLOBULAR PROTEIN HORMONE
6	1QBF	NMR SOLUTION STRUCTURE OF PORCINE PEPTIDE YY
7	1RON	NMR SOLUTION STRUCTURE OF HUMAN NEUROPEPTIDE Y
8	1RU5	SOLUTION STRUCTURE OF PORCINE PEPTIDE YY (PPYY)
9	1RUU	SOLUTION STRUCTURE OF PORCINE PEPTIDE YY (PPYY) BOUND TO DPC MICELLES
10	1TZ4	[HPP19-23]-PNPY BOUND TO DPC MICELLES
11	1TZ5	[PNPY19-23]-HPP BOUND TO DPC MICELLES
12	2DEZ	STRUCTURE OF HUMAN PYY
13	2DF0	SOLUTION STRUCTURE OF HUMAN PYY3-36

On the bases of above results the translated nucleotide sequence ADD83204.1 was found to be a neuropeptide having pancreatic hormone domain and motif pattern in the amino acid sequence. Pancreatic hormone domain is a regulator of pancreatic and gastrointestinal functions and amino acid sequence ADD83204.1 was found to be are closely related with neuropeptide Y (NPY) b, peptide YY (PYY), and pancreatic polypeptide (PP).

REFERENCES

1. BirdLife International (2004). Emberiza bruniceps. 2006. IUCN Red List of Threatened Species. IUCN 2006.

2. Buntings and Sparrows by Byers, Olsson and Curson, ISBN 1-873403-19-4.

3. Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

4. Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

5. Marchler-Bauer A et al. (2009), "CDD: specific functional annotation with the Conserved Domain Database.", Nucleic Acids Res.37(D)205-10.

6. Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", Nucleic Acids Res.32(W)327-331

7. Falquet L., Pagni M., Bucher P., Hulo N., Sigrist C.J, Hofmann K. and Bairoch A. "THE PROSITE database, its status in 2002"Nucl. Acids Res. 30(1):235-238, 2002.

8. Bucher P., Bairoch A. A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation In "ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology."

About Corresponding Author: Mr. Satendra Singh

Mr. Satendra Singh graduated in Pharmacy at UPTU and post graduate in Information Technology at IIIT Allahabad. At post graduate level taken specialization in Bioinformatics. His main area of research includes genome analysis, drug designing, molecular dynamics and chemoinformatics. Currently working as Assistant Professor at the Sam Higginbotom Institute of Agriculture, Technology & Sciences, Deemed University, Allahabad, India.