Peptides 30 (2009) 1241–1248

Contents lists available at ScienceDirect

Peptides

journal homepage: www.elsevier.com/locate/peptides

FGLamide Allatostatin genes in Arthropoda: Introns early or late?

Francisco Martínez-Pérez [a,b], William G. Bendena [c], Belinda S.W. Chang [a], Stephen S. Tobe [a,*]

[a] Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada
[b] Department of Genetics and Molecular Biology, CINVESTAV, Mexico, D.F., Mexico
[c] Department of Biology, Queen's University, Kingston, ON, Canada

ARTICLE INFO

Article history:
Received 28 February 2009
Received in revised form 1 April 2009
Accepted 2 April 2009
Available online 10 April 2009

Keywords:
Allatostatins
Arthropoda
Gene structure
Introns
Late model

ABSTRACT

FGLamide allatostatins are invertebrate neuropeptides which inhibit juvenile hormone biosynthesis in Dictyoptera and related orders and also show myomodulatory activity. The FGLamide allatostatin (AST) gene structure in Dictyoptera is intronless within the ORF, whereas in 9 species of Diptera, the FGLamide AST ORF has one intron. To investigate the evolutionary history of AST intron structure, (intron early versus intron late hypothesis), all available Arthropoda FGLamide AST gene sequences were examined from genome databases with reference to intron presence and position/phase. Three types of FGLamide AST ORF organization were found: intronless in *I. scapularis* and *P. humanus corporis*; one intron in *D. pulex*, *A. pisum*, *A. mellifera* and five *Drosophila* sp.; two introns in *N. vitripennis*, *B. mori* strains, *A. aegypti*, *A. gambiae* and *C. quinquefasciatus*. The literature suggests that for the majority of genes examined, most introns exist between codons (phase 0) which may reflect an ancient function of introns to separate protein modules. 60% of the FGLamide AST ORFs introns were between the first and second base within a codon (phase 1), 28% were between the second and third nucleotides within a codon (phase two) and 12% were phase 0. As would be required for correct intron splicing consensus sequence, 84% of introns were in codons starting with guanine. The positioning of introns was a maximum of 9 codons from a dibasic cleavage site. Our results suggest that the introns in the analyzed species support the intron late model.

1. Introduction

The FGLamide allatostatins (ASTs) are neuropeptides which are potent inhibitors of juvenile hormone (JH) biosynthesis in cockroaches [44], termites [45] and crickets [24], but do not have any known effect on JH production in insects of any other order [35]. They may also serve as neurotransmitters or neuromodulators in the central and peripheral nervous systems, and/or as hormones when released from endocrine cells in the gut [4]. *In vitro* assays have demonstrated that FGLamide ASTs have a potent inhibitory effect on spontaneous contractions of various regions of the blowfly gut [6]. In crustaceans, the FGLamide ASTs can stimulate production of some intermediates in the JH biosynthetic pathway [20]. The FGLamide ASTs have variable amino acid numbers: 6–18 in insects and 6–20 in crustaceans [13]. The N-terminal amino acid sequence is variable, but the C-terminal core sequence Y/FXFGL/I/V-amide is essential for the inhibitory effect on JH biosynthesis, and it is present in all FGLamide ASTs [15]. The FGLamide ASTs are synthesized as a prepropeptide in which the number of FGLamide ASTs processed into mature peptides is species-specific [13,36]. In the precursor, each FGLamide AST is flanked by a C-terminal glycine required for amidation immediately followed by a KR/RR, potential endoproteolytic cleavage site [2,36,39,40]. The intron and exon organization of the FGLamide AST open reading frame (ORF) have been established for insects. In 6 species from Dictyoptera, the FGLamide AST ORF does not have introns [2], whereas in 7 species of Diptera from the genus *Drosophila*, *Calliphora vomitoria* and *Lucilia cuprina*, the gene has two introns [4,7]. These differences in the intron number raise several questions: what were the origins of introns in the FGLamide AST ORF of *Drosophila* species? Do other FGLamide AST ORFs from Arthropoda have introns?

In recent years, two opposite points of view about the origin of introns in eukaryotic cells have been proposed: the intron early hypothesis – the introns are extremely ancient structures, dating to before the divergence between the three eukaryotic kingdoms [23,31,33,34]; the intron late hypothesis – the introns have arisen more recently during early eukaryotic evolution [23,31,33,34]. Understanding of the evolution of spliceosomal introns, with respect to the dynamics of both intron loss and gain, has increased exponentially over the past few years, as a result of the release of many genome sequences of major eukaryotic lineages [14].

* Corresponding author at: Department of Cell and Systems Biology, University of Toronto, 25 Harbord Street, Toronto, ON M5S 3G5, Canada. Tel.: +1 416 978 3517; fax: +1 416 978 3522.
E-mail address: stephen.tobe@utoronto.ca (S.S. Tobe).

0196-9781/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.peptides.2009.04.001

Studies of individual genes and genomes have revealed that some intron positions have been conserved through eukaryotic evolution [14]. The intron position in any ORF can occur in three different positions: phase zero—between two codons, phase one—between the first and second nucleotide of a codon and phase two—between the second and third nucleotide of a codon [23,28]. The position of the introns could be ancient, for example, when these have been conserved in at least two eukaryotic kingdoms (animal, fungi and plants), or in at least two distant animal lineages (e.g. nematodes, arthropods, and vertebrates), whereas an intron can be considered recent if the intron phase and position are specific for an eukaryotic lineage [23,34].

The existence of protosplice sites has been addressed directly by examining the context of introns inserted within codons encoding amino acids that are conserved in all eukaryotes and that, accordingly, are not subjected to selection for splicing efficiency [1,41,43]. Evidence has been presented that introns are either predominantly inserted into specific protosplice sites, which have the exon|intron consensus sequence (A/C)AG|GT, or are inserted randomly but preferentially fixed at such sites [1,41,43]. For example, a preference for phase one introns occurs when a glycine codon (GGN) is localized near codons for dibasic proteolytic sites in secreted proteins [1,41,43]. The aim of the present work was to determine if the introns in the ORF of the FGLamide AST gene are recent or ancient as revealed in the Arthropoda genome databases.

2. Material and methods

2.1. Genome data base

FGLamide AST genes and mRNA were obtained via keyword searches for: *Apis mellifera*, *Nasonia vitripennis*, *Acyrthosiphon pisum*, *Pediculus humanus corporis*, *Bombyx mori* strain p50T, *B. mori* variety Dazao, *Drosophila ananassae*, *Drosophila grimshawi*, *Drosophila melanogaster*, *Drosophila mojavensis*, *Drosophila pseudoobscura*, *Drosophila simulans*, *Drosophila yakuba*, *Drosophila erecta*, *Drosophila persimilis*, *Drosophila sechellia*, *Drosophila virilis*, *Drosophila willistoni*, *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus* and *Ixodes scapularis* (Arachnida) These sequences were obtained from the following databases: National Center for Biotechnology Information, Fly Base [42] and Human Genome Sequencing Center. The FGLamide AST gene and mRNA from *Daphnia pulex* was obtained from Daphnia pulex V1.0 from the Joint Genome Institute.

2.2. In silico determination of FGLamide AST gene and mRNA in genome data base

To determine FGLamide AST gene structure not reported in some genome databases, BLAST GENOME was used with FGLamide allatostatin cDNA from *D. punctata* and *D. melanogaster*. With the sequences obtained, the exon–intron organization was established with Softberry programs (http://linux1.softberry.com/berry.phtml). The conceptual translation and the sequence edition for each gene were done with the software Translate from EXPASY (http://ca.expasy.org/) and Molecular Kit tools (http://www.vivo.colostate.edu/molkit/index.html).

2.3. FGLamide AST gene and mRNA in database construction

For all comparative analysis of FGLamide AST ORF structures, a database was constructed with the sequences obtained and FGLamide AST ORFs reported previously from: *Diploptera punctata*, *Periplaneta americana*, *Blattella germanica*, *Blaberus craniifer*, *Blatta orientalis*, *Supella longipalpa*, *C. vomitoria*, *L. cuprina*, *D. ananassae*, *D. grimshawi*, *D. melanogaster*, *D. mojavensis*, *D. pseudoobscura*, *D. simulans* and *D. yakuba* [2,4,7].

2.4. Phase and intron position

The intron phase from each gene was determined using alignments of the gene and mRNA with: ClustalW [38]. To determine the intron position between the precursors, multiple alignments were built using ClustalW (with gap 0.05 and window 9). The alignment was edited with GENEDOC software [27]. The phylograms were edited with MEGA [37,19].

2.5. Phylogenetic analyses

FGLamide AST precursor proteins were aligned using ClustalW [21,38] and subjected to phylogenetic analyses using maximum likelihood and Bayesian methods [11,32]. Maximum likelihood phylogenetic methods were implemented in the program PHYML 3.0 [11,12] using the LG amino acid replacement matrix [22]. For the likelihood analyses, bootstrapping methods were used to assess the degree of confidence in nodes of the phylogeny [8]. Bayesian inference was performed in MrBayes 3.1.2 [32] using model jumping among fixed-rate amino acid models, with all of the AST sequence data in a single partition. A single Markov chain Monte Carlo run was performed, with four chains (three heated and one cold) for 1 million generations, sampling trees (and parameters) every 1000 generations. Stationarity was assessed using Tracer 1.4 [30] and the first 25% of trees discarded as burn-in. Remaining trees were taken as representative of the posterior probability distribution.

3. Results

3.1. Identification of FGLamide AST ORFs and mRNA in genome databases from Arthropoda

The C-terminus consensus to FGLamide ASTs from Arthropoda was found in all genome databases analyzed. In the majority of sequences, the gene and mRNA were reported by each Genome Project as FGLamide AST; however, in *P. humanus corporis* and five *Drosophila* sp., some of the FGLamide AST ORFs had different names, whereas in *N. vitripennis*, *A. pisum*, *B. mori* strain *p50T* and *B. mori* strain *Davao*, the exon and intron organization were not previously established (Table 1).

3.2. FGLamide AST ORF organization in Arthropoda

Three different forms of FGLamide AST ORF organization were found: (1) the ORF has no introns, and therefore the mRNA for the encoded prepropeptide originated from a single exon. This pattern was found in 2 species: the tick *I. scapularis* in which the precursor has four potential copies of FGLamide ASTs, and the insect *P. humanus corporis* with 6 copies FGLamide AST copies. (2) The second type of FGLamide AST ORF organization shows one intron and two exons. The genera that possess this organization had differing FGLamide AST number. In this gene structure, the first exon codes for the majority of FGLamide ASTs, whereas the second exon gave one or two FGLamide ASTs. This organization was observed in the crustacean *D. pulex* and in the insects *Drosophila* sp., *C. vomitoria*, *L. cuprina* and *A. pisum*, whereas *A. mellifera* had a different organization, with one FGLamide AST in the first exon and the rest in the second exon. (3) In this third type of organization, the ORF has two introns and three exons. This pattern was found in *N. vitripennis*, *B. mori strain p50T*, *B. mori strain Davao*, *A. aegypti*, *A. gambiae* and *C. quinquefasciatus*. The first exon encoded one or two FGLamide ASTs. The second exon possessed codons for the amino terminus for one or more FGLamide ASTs and the last exon had codons for the carboxyl terminus of one or more FGLamide ASTs, depending on the species (Table 2).