General and Comparative Endocrinology 202 (2014) 76–86

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

Diffusible gas transmitter signaling in the copepod crustacean *Calanus finmarchicus*: Identification of the biosynthetic enzymes of nitric oxide (NO), carbon monoxide (CO) and hydrogen sulfide (H_2S) using a *de novo* assembled transcriptome

Andrew E. Christie *, Tiana M. Fontanilla, Vittoria Roncalli, Matthew C. Cieslak, Petra H. Lenz

Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA

ARTICLE INFO

Article history:
Received 27 January 2014
Revised 26 March 2014
Accepted 1 April 2014
Available online 18 April 2014

Keywords:
Nitric oxide synthase (NOS)
Heme oxygenase (HO)
Cystathionine β-synthase (CBS)
Illumina sequencing
Neurotransmitter
Neuromodulator

ABSTRACT

Neurochemical signaling is a major component of physiological/behavioral control throughout the animal kingdom. Gas transmitters are perhaps the most ancient class of molecules used by nervous systems for chemical communication. Three gases are generally recognized as being produced by neurons: nitric oxide (NO), carbon monoxide (CO) and hydrogen sulfide (H_2S). As part of an ongoing effort to identify and characterize the neurochemical signaling systems of the copepod *Calanus finmarchicus*, the biomass dominant zooplankton in much of the North Atlantic Ocean, we have mined a *de novo* assembled transcriptome for sequences encoding the neuronal biosynthetic enzymes of these gases, *i.e.* nitric oxide synthase (NOS), heme oxygenase (HO) and cystathionine β-synthase (CBS), respectively. Using *Drosophila* proteins as queries, two NOS-, one HO-, and one CBS-encoding transcripts were identified. Reverse BLAST and structural analyses of the deduced proteins suggest that each is a true member of its respective enzyme family. RNA-Seq data collected from embryos, early nauplii, late nauplii, early copepodites, late copepodites and adults revealed the expression of each transcript to be stage specific: one NOS restricted primarily to the embryo and the other was absent in the embryo but expressed in all other stages, no CBS expression in the embryo, but present in all other stages, and HO expressed across all developmental stages. Given the importance of gas transmitters in the regulatory control of a number of physiological processes, these data open opportunities for investigating the roles these proteins play under different life-stage and environmental conditions in this ecologically important species.

1. Introduction

The molecules used for neurochemical signaling are large in number and diverse in structure, but can be divided into several broad categories including peptides, amines, small molecule transmitters and gas transmitters. The latter class is considered among the most ancient of these signaling molecules (*e.g.* Garthwaite, 2008), with diffusible gases being used for chemical communication in essentially all prokaryotic and eukaryotic organisms (*e.g.* Crane et al., 2010; Lloyd, 2006; Sudhamsu and Crane, 2009; Wang, 2012; Wilson et al., 2012). In nervous systems, three diffusible gases, nitric oxide (NO), carbon monoxide (CO) and hydrogen sulfide (H_2S), are now generally recognized as serving roles as neurotransmitters/modulators (*e.g.* Althaus and Clauss, 2013; Barañano et al., 2001; Barañano and Snyder, 2001; Boehning and Snyder, 2003; Gadalla and Snyder, 2010; Garthwaite, 2008; Mustafa et al., 2009; Snyder and Ferris, 2000; Vandiver and Snyder, 2012).

Signaling by diffusible gases has been described as "atypical" with respect to the other commonly recognized classes of neurochemicals. Specifically, unlike peptides, amines and small molecule transmitters, gas transmitters are not stored in vesicles and are not released via exocytosis (*e.g.* Snyder and Ferris, 2000). Instead, NO, CO and H_2S are synthesized as needed, and are freely diffusible and membrane permeant once produced. Thus, the sphere of influence of a gas transmitter is not constrained by physical barriers such as desmosomes and other junctional complexes, but is limited only by its diffusion constant and life span. Given these factors, the control of gas transmitters relies largely on the fine tuned regulation of their biosynthetic enzymes (*e.g.* Mustafa et al., 2009), *i.e.* nitric oxide synthase (NOS) for NO, heme oxygenase (HO) for CO, and cystathionine β-synthase (CBS) for H_2S (*e.g.* Althaus and Clauss, 2013).

* Corresponding author. Fax: +1 808 956 6984.
E-mail address: crabman@pbrc.hawaii.edu (A.E. Christie).

http://dx.doi.org/10.1016/j.ygcen.2014.04.003
0016-6480/

As part of an ongoing effort to identify and characterize the neurochemical signaling systems of the copepod crustacean *Calanus finmarchicus*, the biomass dominant zooplankton species for much of the North Atlantic Ocean (Dale et al., 2001; Head et al., 2000; Marshall and Orr, 1955; Meise and O'Reilly, 1996), we have mined a *de novo* assembled transcriptome (Christie et al., 2013a; Lenz et al., 2014) for sequences encoding putative gas transmitter biosynthetic enzymes. In addition, the proteins deduced from the identified transcripts were assessed for structural motifs characteristic of the respective enzyme family. Finally, RNA-Seq data collected from six *C. finmarchicus* developmental stages (embryo, early nauplius, late nauplius, early copepodite, late copepodite and adult) were used to map the developmental expression of the enzymes in this species. As our data will show, NOS-, HO-, and CBS-encoding transcripts were identified from our transcriptome assembly; for NOS, two different sequences were found. The proteins deduced from the identified transcripts all possess structural features characteristic of their respective family, and reverse BLAST analyses support these enzyme family attributions. Expression mapping of the two NOS-encoding transcripts revealed one to be expressed in embryos and essentially absent in all other stages, and the other to be absent in embryos, but present in all other developmental stages. The CBS-encoding sequence was absent in embryo but expressed in all other stages, while the HO-encoding transcript was expressed throughout development. These data represent the first descriptions of gas transmitter biosynthetic enzymes in *C. finmarchicus*, suggest that these proteins are stage-specific in their expression, and lay a strong foundation for future anatomical, biochemical and physiological investigations of diffusible gas signaling in this environmentally critical copepod species.

2. Materials and methods

2.1. De novo transcriptome assembly

A *de novo* transcriptome for *C. finmarchicus* was generated as described in detail in Lenz et al. (2014). In brief, multiplexed gene libraries were prepared from RNA extracted from six developmental stages of wild-caught or laboratory reared *C. finmarchicus*. The developmental stages included in the libraries were egg (which represents a mixture of embryonic stages), early nauplius (stages NI and NII), late nauplius (stages NV and NVI), early copepodite (stages CI and CII), late copepodite (stage CV) and adult female; all libraries were sequenced at the HudsonAlpha Institute for Biotechnology (Huntsville, AL, USA) in a single lane using an Illumina HiSeq 2000 instrument (Illumina Inc.). In total, 415,469,690 raw, 100 bp, paired-end reads were obtained. After quality filtering and trimming, the combined raw reads from all six developmental stage libraries were assembled *de novo* using Trinity 2012-03-17-IU_ZIH_TUNED software on a node of the National Center for Genome Analysis Support's (NCGAS; Indiana University, Bloomington, IN, USA) Mason Linux cluster. In total, 206,041 unique nucleotide sequences were generated using Trinity.

2.2. Transcriptome mining

Searches of the transcriptome assembly produced by Trinity were conducted using the DeCypher Tera-BLASTP algorithm on the Mount Desert Island Biological Laboratory's TimeLogic DeCypher server (MDIBL, Salisbury Cove, ME, USA; http://decypher.mdibl.org/decypher/algo-tera-blast/tera-tblastn_an.shtml) as described previously (Christie et al., 2013a,b, 2014). For all searches, the DeCypher program database was set to "6libTrinity", the combined Trinity assembly, and a known fruit fly *Drosophila melanogaster* protein was used as the query. All hits were translated and checked manually for homology to the target query. Table 1 provides the BLAST-generated *E*-value for each hit that was identified as encoding a putative target transcript, as well as the lengths of identified transcripts; the length of the protein deduced from each target sequence is also provided in this table.

2.3. Analyses of protein conservation and structure

Analyses of protein conservation and structure were conducted using a protocol described in detail in several recent publications (*e.g.* Christie et al., 2013a,b, 2014). To identify the most similar arthropod proteins to each of the *C. finmarchicus* enzymes described in this study, the deduced *Calanus* sequence was used to query the non-redundant arthropod protein dataset (taxid:6656) curated in GenBank (excluding *C. finmarchicus* proteins, obvious partial proteins and synthetic constructs) using the blastp algorithm (Altschul et al., 1997); the results of these searches are summarized in Table 2. In addition, the identified *Calanus* NOS and HO isoforms were used as queries against selected vertebrate protein databases (*i.e.* pufferfish *Takifugu rubripes* [taxid:31033], the clawed frog *Xenopus laevis* [taxid:8355], the chicken *Gallus gallus* [taxid:9031], and human *Homo sapiens* [taxid: 9606]) in order to determine the subfamilies into which they likely reside; these data are summarized in Table 3. All database searches were conducted on or before May 20, 2013.

To determine amino acid identity/similarity between proteins, the sequences in question were aligned using MAFFT version 7 (http://align.bmr.kyushu-u.ak.jp/mafft/online/server/; Katoh and Standley, 2013), and amino acid identity/similarity was subsequently determined using the alignment output. Percent identity was calculated as number of identical amino acids (denoted by "*" in the MAFFT output) divided by the total number of amino acids in the longest sequence (×100). Percent similarity was calculated as number of identical and similar amino acids (the latter denoted by the ":" and "." symbols in the protein alignment) divided by the total number of amino acids in longest sequence (×100).

Protein structural motifs were analyzed using the online program Pfam version 27.0 (http://pfam.sanger.ac.uk/; Punta et al., 2012). A common highlighting scheme has been used to denote functional domains in the relevant figures: nitric oxide synthase oxygenase domain, green; flavodoxin domain, dark blue; FAD-binding domain, light blue; oxidoreductase NAD-binding domain, black; heme oxygenase domain, red; pyridoxal-phosphate dependent enzyme domain, yellow; cystathionine β-synthase domain, pink.

2.4. Vetting of deduced protein sequences using publicly accessible expressed sequence tags

In an attempt to confirm the amino acid sequences of the proteins deduced from the Trinity assembly, each of the putative *Calanus* proteins was used as a query to search the extant *C. finmarchicus* ESTs (~11,000 in total; Lenz et al., 2012) curated at NCBI using the tblastn algorithm as described previously (Christie et al., 2013a,b, 2014).

2.5. Developmental expression mapping

To compare the relative levels of transcript expression in embryos, early nauplii, late nauplii, early copepodites, late copepodites, and adult females, Illumina reads from each of these developmental stages were mapped against the sequence in question using Bowtie software (Johns Hopkins University, Baltimore, MD, USA; http://bowtie-bio.sourceforge.net/index.shtml; Langmead et al., 2009) as described previously (Christie et al., 2013a,c,