Developmental and Comparative Immunology 33 (2009) 881–889

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

The diversity of C-type lectins in the genome of a basal metazoan, *Nematostella vectensis*

Elisha M. Wood-Charlson [a,b,*], Virginia M. Weis [b]

[a] Department of Oceanography, University of Hawai'i at Manoa, Honolulu, HI 96822, United States
[b] Department of Zoology, Oregon State University, Corvallis, OR 97331, United States

ARTICLE INFO

Article history:
Received 16 December 2008
Received in revised form 29 January 2009
Accepted 30 January 2009
Available online 21 February 2009

Keywords:
Cnidarian
C-type
Glycan
Innate immunity
Lectin
Nematostella
Symbiosis

ABSTRACT

C-type lectins (CTLs) are involved in cell–cell adhesion, recognition, and innate immunity in higher vertebrates, but little is known about CTLs in basal metazoans. The recent sequencing of the cnidarian *Nematostella vectensis* genome allowed us to explore the CTL-like gene family at the base of metazoan evolution. Sixty-seven predicted CTLs, with a total of 92 putative C-type lectin domains (CTLDs), were classified according to number of CTLDs present and their association with other protein domains in the CTL. Conserved residues in the glycan-binding pocket suggest that approximately half of the CTLDs retain glycan-binding function. Phylogenetic analysis of *N. vectensis* CTLDs with respect to other model invertebrates and humans indicates *N. vectensis* CTLD sequences more closely resemble vertebrate CTLDs. This study provides a *N. vectensis* CTL database that can be used for further research on the evolution of cnidarian CTLs and the role of CTLs in cnidarian innate immunity.

1. Introduction

The recent expansion of genome sequencing and the generation of large EST datasets is beginning to allow for studies on the evolution of gene families. Many of these studies have focused on gene families that have been conserved throughout metazoan evolution [1–5]. Most of the major gene families involved in conserved, developmentally regulated, vertebrate signaling pathways have also been identified in cnidarians, a basal metazoan group [4,6] (Fig. 1). Recently, similar comparisons have begun for less well-conserved gene families, such as receptors that function in innate immunity. These immunity gene families, which are also present in cnidarians [7], are likely involved in "trench warfare"—continuous, diversifying selection pressure imposed by host–microbe interactions [8,9]. One example of an innate immune receptor family is the C-type lectin (CTL) gene family, which has been described for model vertebrate and invertebrate organisms [10–13], but has not been investigated for basal metazoans.

CTLs make up a very diverse gene family that is identified by a C-type lectin-like domain (CTLD), which bind glycans in a Ca^{2+}-dependent manner. The CTLD ranges from 115 to 130 amino acids in length and can be identified by 14 invariant and 18 highly conserved amino acids [14]. Most of the conserved residues occur within two functional sites: the first binds a single Ca^{2+} ion and the second binds another Ca^{2+} and the glycan ligand [15] (Fig. 2). Outside of these motifs, CTLD sequences are not well conserved, and CTLs often contain more than one CTLD in addition to numerous other domains that determine many of the CTL functions [16].

CTLDs play a role in a variety of biological events that require recognition of specific glycans, such as cell–cell adhesion, recognition, and phagocytosis of potential pathogens [17–19]. Since cell–cell adhesion and immunity have been major contributors to metazoan evolution [20,21], understanding the diversity within the CTL gene family will provide valuable insight into ancestral metazoan complexity. Analyses of CTLDs from two invertebrate model organisms, *Drosophila* and *Caenorhabditis*, suggested that CTLs diverged dramatically since the split between vertebrate and invertebrate lineages [22,23]. However, no study has examined the diversity of the CTL gene family in basal metazoans, such as cnidarians, which harbor a surprising amount

* Corresponding author at: 1000 Pope Road, Marine Sciences Building, Honolulu, HI 96822, United States. Tel.: +1 808 956 7633; fax: +1 808 956 9225.
E-mail address: woodchae@lifetime.oregonstate.edu (E.M. Wood-Charlson).

Abbreviations: CTL, C-type lectin; CTLD, C-type lectin-like domain; EGF, epidermal growth factor; F58C, coagulation factor 5/8 type C; Gal, galactose; Gal-lectin, D-galactoside-binding lectin; Glc, glucose; GPS, G-protein coupled receptor (GPCR) proteolytic site; Ig, immunoglobulin-like; LDLRA, low-density lipoprotein receptor A; Man, mannose; SCP, sperm-coating glycoprotein; 7tm, 7-pass transmembrane GPCR; Signal, signal sequence; SRCR, scavenger receptor cysteine rich; tm, transmembrane domain; TSP1, thrombospondin type I repeat; vWA, von Willebrand factor A.

0145-305X/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2009.01.008

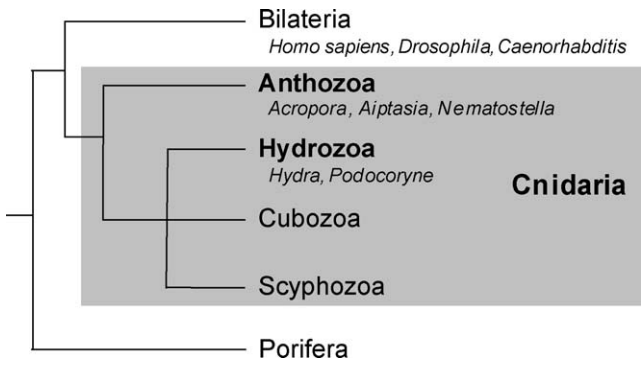

Fig. 1. Cnidarians as a basal group in metazoan evolution. Within the Cnidaria, the Anthozoa are basal and the Hydrozoa are derived (modified from [7]).

of genomic complexity that, in many cases, more closely reflects vertebrate-level complexity than the other invertebrate model organisms [2–4,24,25]. Aside from preliminary genomics work, there are a few studies that have explored the functional role of lectins in cnidarians. Cnidarians, such as anemones and corals, are often found in a mutualistic relationship with a single-celled dinoflagellate. In most cases, these very selective relationships must be established anew for every host generation and requires a complex series of steps. These steps include recognition between the appropriate symbiotic partners and phagocytosis of the symbiont by the host (reviewed in [26,27])—both known functions of lectin/glycan interactions. During the onset of symbiosis, lectin/glycan interactions appear to provide a mechanism of recognition between host cnidarians and their algal symbionts [28–30]. There is also evidence that cnidarian lectins can harbor extensive sequence variation and be used to bind potential bacterial pathogens, as well as potential symbionts [31]. The identity, localization, and function of cnidarian lectins during other stages in the symbiont selection process or other aspects of innate immunity remain a mystery.

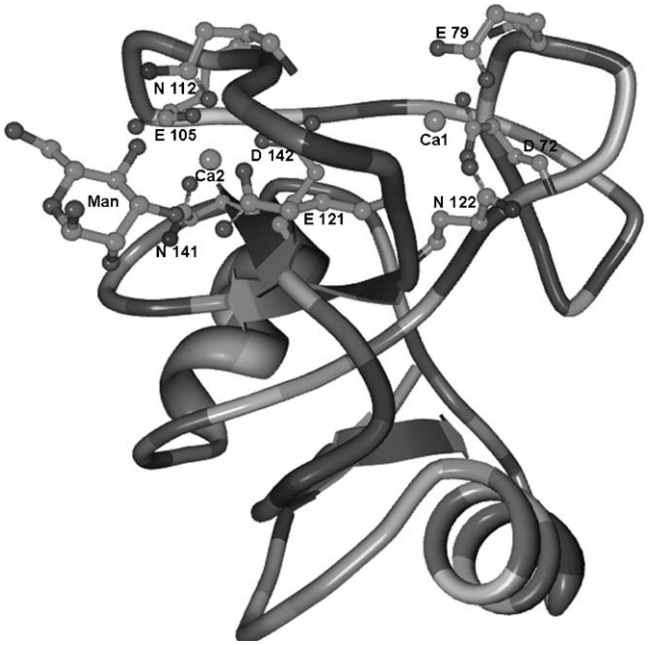

Fig. 2. Structure of a typical CTLD in complex with a glycan ligand (generated from Protein Data Bank: 1kzc, www.rcsb.org/pdb). The residues that create the two Ca^{2+}-binding pockets are identified by their single-letter amino acid abbreviation and sequence position that corresponds to the alignment in Fig. 3. The second Ca^{2+}-binding pocket also binds the glycan ligand (shown here in association with Man).

Cnidarians, such as the non-symbiotic sea anemone *Nematostella vectensis*, represent a sister-group to the Bilateria (Fig. 1). *N. vectensis* has become a cnidarian model for developmental studies, and the recent sequencing of its genome allows for analyses of gene families in a basal metazoan [2]. This study identifies and describes members of the CTL gene family in *N. vectensis* for future use in evolutionary and functional innate immunity studies.

2. Materials and methods

2.1. Database searching

Using the Joint Genome Institute (JGI) interactive *N. vectensis* genome browser (v1.0) (http://genome.jgi-psf.org/Nemve1/Nemve1.home.html), searches were conducted to retrieve potential CTL-like sequences from the genome. A tblastn search ($e < 10^{-5}$) was performed using CTLD sequences from human, mouse, and 13 invertebrates (query sequences are available by request). In addition, the annotated *N. vectensis* genome was searched for key terms: lectin (280 hits), C-type (201 hits), mannose (18 hits), galactose (16 hits), and InterPro C-type lectin (IPR001304, 89 hits). CTLD-containing sequences from other cnidarians were identified by a tblastn search ($e < 10^{-5}$) of the cnidarian sequences (taxonomy ID: 6073) at the National Center for Biotechnology Information (NCBI), the *Hydra magnipapillata* genome browser (v1.64, http://hydrazome.metazome.net/cgibin/gbrowse/hydra/), and an EST database from *Aiptasia pallida* (http://aiptasia.cs.vassar.edu/AiptasiaBase/index.php) [32].

2.2. Assessing gene prediction models

For all CTLD-containing sequences, the coding region was identified by the *Ab initio* model using Fgenesh gene prediction (available on the JGI genome browser) and only modified when EST data were available. In some cases, the *Ab initio* model may have missed a potential CTLD coding sequence and/or not predicted the full-length CTLD; however, the *Ab initio* model was selected for consistency because it predicted a gene model for all putative CTLD sequences retrieved from the database searches. Using a combination of methods, the predicted CTL sequences were screened to confirm the presence of at least one CTLD: (1) blast search against the NCBI protein database resulted in a significant hit to a CTLD-containing sequence ($e < 10^{-5}$) and (2) protein domain search against PfamA [33], ScanProsite [34], and InterProScan [35] resulted in a CTLD profile hit from more than one database. The complete list of CTLD-containing gene predictions from the *N. vectensis* genome can be found in Table 1 and viewed on the JGI genome browser by searching Gene Models with the gene model name as listed.

2.3. Identification of domains within gene predictions

Using the compiled dataset of *N. vectensis* CTL-like sequences, additional protein domains within the predicted coding regions were identified by PfamA, ScanProsite, and InterProScan. Domains were included if they were predicted by more than one annotation method or if a single annotation method had a significant hit to the domain ($e < 10^{-5}$). Putative signal sequences and transmembrane domains were identified by InterProScan and confirmed by hydropathy plots.

2.4. Sequence analysis

CTLDs were trimmed at the longest disulfide bond and aligned using ClustalX [36]. CTLDs have been predicted to have functional glycan-binding if they contain at least three out of the five conserved residues known to function in the Ca^{2+}/glycan-binding