Genomics 99 (2012) 282–291

Contents lists available at SciVerse ScienceDirect

Genomics

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

Genomic and functional characterization of the diverse immunoglobulin domain-containing protein (DICP) family ☆

Robert N. Haire [a], John P. Cannon [a], Marci L. O'Driscoll [b], David A. Ostrov [c], M. Gail Mueller [a,b], Poem M. Turner [d], Ronda T. Litman [a], Gary W. Litman [a,b,e], Jeffrey A. Yoder [d,*]

[a] Department of Pediatrics, University of South Florida College of Medicine, USF/ACH Children's Research Institute, 140 7th Avenue South, St. Petersburg, FL 33701, USA
[b] Department of Molecular Genetics, All Children's Hospital, 501 6th Avenue South, St. Petersburg, FL 33701, USA
[c] Department of Pathology, Immunology and Laboratory Medicine, University of Florida College of Medicine, Gainesville, FL 32610, USA
[d] Department of Molecular Biomedical Sciences and Center for Comparative Medicine and Translational Research, College of Veterinary Medicine, North Carolina State University, 1060 William Moore Drive, Raleigh, NC 27607, USA
[e] H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive, Tampa, FL 33612, USA

ARTICLE INFO

Article history:
Received 30 October 2011
Accepted 10 February 2012
Available online 21 February 2012

Keywords:
Zebrafish
Innate immunity
Lipid binding

ABSTRACT

A heretofore-unrecognized multigene family encoding diverse immunoglobulin (Ig) domain-containing proteins (DICPs) was identified in the zebrafish genome. Twenty-nine distinct loci mapping to three chromosomal regions encode receptor-type structures possessing two classes of Ig ectodomains (D1 and D2). The sequence and number of Ig domains, transmembrane regions and signaling motifs vary between DICPs. Interindividual polymorphism and alternative RNA processing contribute to DICP diversity. Molecular models indicate that most D1 domains are of the variable (V) type; D2 domains are Ig-like. Sequence differences between D1 domains are concentrated in hypervariable regions on the front sheet strands of the Ig fold. Recombinant DICP Ig domains bind lipids, a property shared by mammalian CD300 and TREM family members. These findings suggest that novel multigene families encoding diversified immune receptors have arisen in different vertebrate lineages and affect parallel patterns of ligand recognition that potentially impact species-specific advantages.

1. Introduction

As the phylogenetically widely divergent species in which immune receptors have been characterized increases, several major trends can be recognized: 1) innate immune receptors have a long evolutionary history with marked similarities in receptor structure and function across wide phylogenetic boundaries [1], 2) primary mediators of adaptive immunity have undergone many changes during the evolution of vertebrates but share remarkable similarities in basic aspects of genetic recombination (rearrangement) and clonal selection [2] and 3) structures of receptors that mediate natural killer (NK)-type function can vary markedly even within members of a single class of vertebrate species (mammals) [3]. It is more difficult to recognize common features of other receptors that are classified as immune-type because of their structural domain composition and signaling properties. Many of these genes are encoded in multigene families and exhibit patterns of structural variation that are predicted to be associated with functional differences. It is likely that at least some receptors encoded by these genes are elements of unrecognized receptor-signaling networks and function through novel mechanisms. The presence of such multigene families in modern representatives of phylogenetically important species emphasizes their significance. Of the various nonmammalian animal models in which these molecules have been identified, the zebrafish (*Danio rerio*) offers many unique methodological advantages.

We have described variable (V) region-containing transmembrane receptors (novel immune-type receptors [NITRs]) in zebrafish and other bony fish [4]. NITRs are the most complex family of V region-containing immune-type receptors described thus far outside of immunoglobulin (Ig) and T cell antigen receptors (TCRs) [5]. NITRs function in allogeneic recognition in a manner akin to activating/inhibitory NK receptors [6]. A direct cloning strategy [7] identified a distantly related multigene family (modular domain immune-type receptors [MDIRs]) [8]. Through genome scanning utilizing MDIR and NITR Ig domain sequences, an additional multigene family encoding diverse Ig domain-containing proteins (DICPs) was identified. We describe herein the genomic organization, sequence complexity and predicted protein structures of the DICPs in zebrafish, which likely are unique to bony fish. We also demonstrate that recombinant forms of zebrafish DICP Ig domains bind lipids, which is a shared characteristic with members of the mammalian CD300 and TREM families of innate immune receptors [9,10].

☆ Data deposition: Sequence data have been deposited with GenBank under accession numbers: JN416849–JN416885.

* Corresponding author. Fax: +1 919 513 7301.
E-mail address: Jeff_Yoder@ncsu.edu (J.A. Yoder).

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.02.004

2. Materials and methods

2.1. Bioinformatics

Genomic sequences encoding candidate DICP Ig domains were identified on zebrafish chromosomes 3, 14 and 16 with BLAST searches using MDIR and NITR sequences as queries. In silico translation of each Ig domain indicates that several genes encode a frame shift or premature stop codon, permitting their classification as pseudogenes (Supplemental Materials and methods). Protein sequences were aligned by Clustal W [11]. Phylogenetic trees were constructed from pairwise Poisson correction distances with 2000 bootstrap replications by MEGA5 software [12]. Protein sequence domains were identified with SMART software [13].

2.2. DICP transcripts and genes

A small number of DICP ESTs were identified using BLAST searches of the zebrafish EST database and those appearing to encode full-length proteins were sequenced (Supplemental Materials and methods). Additional DICP cDNA sequences were obtained by rapid amplification of cDNA ends (RACE) or direct reverse transcriptase-polymerase chain reaction (RT-PCR) with primers complementing predicted exons (Supplemental Materials and methods).

2.3. DICP D1–D2 cDNA amplicons from chromosome 3

Partial DICP cDNA sequences were generated using primers designed to amplify D1–D2-containing DICP genes on chromosome 3. Forward (CATGTGTTCAGCAGWTMTGGAGAAACTG) and reverse (GATAGACTCCACATCTCCACTGTTTATC) primers were used with Titanium Taq (Clontech) to amplify D1–D2 sequences from pooled kidney and intestine cDNA (zebrafish obtained from EkkWill Waterlife Resources, Ruskin, FL, USA). Amplicons were cloned into pGEM-T Easy (Promega) and sequenced.

2.4. Genomic organization

The genomic organization of DICPs was deduced by comparing cDNA sequences to ZV8 genomic reference sequences: chromosome 3 scaffold 262 (GenBank ID: NW_001878770.2), chromosome 14 scaffold 1719 (GenBank ID: NW_001877436.2) and chromosome 16 scaffold 1952 (GenBank ID: NW_001877662.2). BACs CH73-34H11 (GenBank ID: FP929011) and CH73-322B17 (GenBank ID: FP015862) were used to link two unordered segments within scaffold 1952 that map to chromosome 16.

2.5. Molecular modeling

Theoretical models of DICP D1 domains were generated using the automated protein homology-modeling server SWISS-MODEL [14]. The Structural Classification of Proteins (SCOP) database was utilized for domain definitions [15]. The Docker program was used to calculate sequence similarity using the Blosum62 matrix. Figures were generated with PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

2.6. Cloning and expression of hFc chimeras

Recombinant soluble proteins of DICP D1 and D2 ectodomains fused to a human IgG Fc domain were generated by cloning various ectodomains (amplified from pooled hematopoietic tissue cDNA) into the pcDNA3-hsIgG1Fc-Avi fusion vector [16] that introduces a N-terminal start codon, signal peptide and a C-terminal human IgG Fc domain.

DICP D1-hFc and D2-hFc chimeric proteins were expressed and secreted by HEK293T cells. Cells were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS), 1 mM sodium pyruvate and 2 mM GlutaMAX (Invitrogen) and transferred to OPTI-MEM I serum-free medium (Invitrogen) for transfection of hFc constructs with Lipofectamine 2000 (Invitrogen). Following transfection, cells were grown for 48 h, pooled, and centrifuged at $500\times g$ for 10 min to clear the supernatant. Recovered supernatants were stored at 4 °C in 0.02% sodium azide. Supernatant harvests were concentrated 10 to 100 fold and the hFc fusion proteins were characterized by Western analyses and quantified using the Easy-Titer Human IgG Assay kit (Thermo Scientific) [16].

2.7. ELISA assay for binding to lipids

Purified lipids (Sigma and Avanti Polar Lipids) were processed as described [9]. Solid phase ELISA assays were conducted as described previously [9]. Either 0.5 μg purified lipid or 50 μl of MBTE/methanol bacterial extract was used to coat plates. Negative control wells were treated in parallel with solvent (100% methanol). Binding efficiency was determined after color development as absorbance at 450 nm. Values were corrected by subtracting the value from negative control wells.

The effect of concentration on lipid binding of hFc fusion proteins in the ELISA assay was evaluated. As a positive control, a hFc-fusion of the Ig domain of murine CLM7, which binds all four purified lipids used in screening [9], was employed. CLM7-hFc was added to ELISA plates at 100 μg/ml (volume 0.10 ml). Dicp1.5^{529}-D1-hFc, which exhibits robust lipid binding, was added at 15 μg/ml (volume 0.10 ml). The optimal lipid binding exhibited by CLM7-hFc was obtained at 12–25 μg/ml [9] and assay results were comparable to that of Dicp1.5^{529}-D1-hFc at 15 μg/ml. The standard concentration of hFc fusion proteins for assays was 0.10 ml of 10–50 μg/ml.

3. Results and discussion

3.1. Identification of DICP Ig domains

A number of approaches exist for identifying immune receptors in diverse species. We employed a robust series of Ig V-, I- and C2-type motifs from NITRs and MDIRs as queries in tBLASTn searches of the zebrafish genome (version Zv8) to identify unrecognized Ig-region encoding genes and identified the DICP family. The typical DICP consists of two distinct classes of extracellular Ig domains: N-terminal D1 and C-terminal D2 domains, (Figs. 1A–C, Supplemental Figs. S1–S2). DICP D1 domains share more conserved residues with classical V domains (G^{16}, V^{19}, L^{21}, C^{23}, W^{41}, L^{89}, I^{91}, D^{98}, G^{100}, Y^{102}, C^{104}) than do the D2 domains (G^{16}, L^{21}, C^{23}, W^{41}, L^{89}, C^{104}) [17]. Additional pairs of conserved cysteines: C^{30} and C^{87} in D1 and C^{33} and C^{85} in D2 (Figs. 1A–B) are predicted to form intrachain disulfides. Twenty-nine DICP D1 domains were identified on zebrafish chromosomes 3, 14 and 16 (Fig. 1D). The genes corresponding to the D1 domains are designated by: a number that denotes the DICP cluster, a number that denotes the order in which the domains were identified and a superscript that indicates an allele sequence source, e.g., *dicp1.7*^{262}: first cluster, seventh D1 domain and scaffold 262.

3.2. DICP transcripts

The sequencing of multiple DICP ESTs and cDNAs (Supplemental Materials and methods and Supplemental Fig. S3) facilitated the characterization of the exon organization and putative translation products from a large number of highly related candidate DICP genes (Fig. 2). Most DICP D1 domain exons are flanked by exons that encode a leader signal sequence and a D2 domain exon; *dicp2.1* and *dicp3.1*