Developmental and Comparative Immunology 35 (2011) 461–468

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

TIR domain-containing adaptor SARM is a late addition to the ongoing microbe–host dialog

Qing Zhang[a,1], Christian M. Zmasek[a], Xiaohui Cai[a], Adam Godzik[a,b,*]

[a] Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
[b] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 26 July 2010
Received in revised form
18 November 2010
Accepted 18 November 2010
Available online 24 November 2010

Keywords:
Toll-like receptor
Host–commensal interaction
Host–pathogen interaction
Lateral gene transfer
Commensal microflora
Innate immunity

ABSTRACT

Toll/interleukin-1 receptor (TIR) domain-containing proteins play important roles in defense against pathogens in both animals and plants, connecting the immunity signaling pathways via a chain of specific protein–protein interactions. Among them is SARM, the only TIR domain-containing adaptor that can negatively regulate TLR signaling. By extensive phylogenetic analysis, we show here that SARM is closely related to bacterial proteins with TIR domains, suggesting that this family has a different evolutionary history from other animal TIR-containing adaptors, possibly emerging via a lateral gene transfer from bacteria to animals. We also show evidence of several similar, independent transfer events, none of which, however, survived in vertebrates. An evolutionary relationship between the animal SARM adaptor and bacterial proteins with TIR domains illustrates the possible role that bacterial TIR-containing proteins play in regulating eukaryotic immune responses and how this mechanism was possibly adapted by the eukaryotes themselves.

1. Introduction

In humans, Toll-like receptors (TLRs) play a critical role in defense against microbial infection, initiating the host innate immune response upon binding microbe-derived molecules (Medzhitov, 2007; West et al., 2006). TLRs use a C-terminal TIR protein–protein interaction domain to connect to downstream adaptor molecules. The TIR domain is also present in the interleukin-1 receptor (IL-1R) family, where it plays a similar role. The signal is carried further by TIR-domain containing adaptors, with receptor TIR domains interacting directly with adaptor TIR domains. Five TIR-containing adaptors have been discovered in human, including four positive regulators—myeloid differentiation factor 88 (MyD88), TIR domain-containing adaptor protein (TIRAP, also known as MAL), TIR domain-containing adaptor inducing interferon-β (TRIF), TRIF-related adaptor molecule (TRAM)—and one negative regulator—sterile α and HEAT-Armadillo motifs containing protein (SARM) (O'Neill and Bowie, 2007). TIR domains are also present in plant pathogen resistance (R) proteins, which elicit hypersensitive responses against pathogen infection, but it is not clear what their downstream targets are (Burch-Smith and Dinesh-Kumar, 2007).

TIR domain-containing proteins have been also found in bacteria (Koonin and Aravind, 2002) and, recently, several of them, mostly from pathogenic bacteria, have been studied in detail. TlpA from *Salmonella enterica serovar enteritidis* (Newman et al., 2006), TcpC from uropathogenic *Escherichia coli* strain CFT073 (Cirl et al., 2008), TcpB from *Brucella melitensis* 16 M (Cirl et al., 2008), and Btp1 from *B. melitensis biovar Abortus 2308* (Salcedo et al., 2008) can all interfere with the host TLR signaling, thus forming a new class of virulence factors. A number of questions arise from these observations. How widely is this interfering mechanism used by bacteria? Is it strictly a virulence mechanism or could it represent a more general mechanism of interaction between the plant or animal hosts and their indigenous microbiota? What is the relationship between TIR-containing proteins from different kingdoms? To address these questions, we searched the NCBI non-redundant (nr) protein database and various metagenomic protein datasets for homologs of well-characterized animal TIR domains and performed a detailed phylogenetic analysis of the TIR family in all kingdoms of life, as well as a structural comparison of the recently solved representatives. Our results show that the TIR family had a complicated evolutionary history that included several independent bacteria-eukaryotes lateral gene transfer events, and moreover, our findings suggest that bacterial TIR domain-containing proteins may play important roles in interactions between bacteria and eukaryotes.

* Corresponding author. Tel.: +1 858 646 3168; fax: +1 858 795 5249.
E-mail address: adam@sanfordburnham.org (A. Godzik).
[1] Current address: Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences, 4301 West Markham Street, Little Rock, AR 72205, USA.

0145-305X/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2010.11.013

2. Materials and methods

2.1. Database search and sequence analysis

The non-redundant (nr) protein database was downloaded from the NCBI FTP site. TIR domain-containing proteins were identified using a cascade of PSI-BLAST searches (Altschul et al., 1997) against the nr protein database by using known human, *Drosophila*, or *C. elegans* TIR domain amino acid sequences as seeds. Up to five iterations of PSI-BLAST were run, and proteins with an *e*-value below 0.005 were taken for further consideration. The selected proteins were then checked by reciprocal BLAST analysis, Pfam protein searches (Bateman et al., 2004), and Conserved Domain Search (CD-Search) (Marchler-Bauer and Bryant, 2004).

Similar strategies were used to identify the TIR domain-containing proteins from human gut (Gill et al., 2006; Kurokawa et al., 2007), soil (Tringe et al., 2005), and global ocean (Venter et al., 2004; Rusch et al., 2007) metagenomic datasets. Taxonomic assignment of the retrieved TIR-containing metagenomic proteins was performed according to the best-hit pairs in the BLASTP analysis against the NCBI nr protein database. To tentatively assign the taxonomic origins of the metagenomic proteins, we adopted a 90% BLASTP identity threshold and the hit length coverage should be more than 60% of the query sequence.

The SEED database (Overbeek et al., 2005) was used to check the gene neighborhood information for bacterial proteins identified in this analysis.

2.2. Multiple sequence alignments, phylogeny reconstructions, and cluster analysis

The extent of the TIR domain for the purpose of multiple alignments was defined by the Pfam 21.0 model of the TIR domain (Bateman et al., 2004). Multiple sequence alignments were produced by MAFFT 6.240 (localpair, maxiterate 1000) (Katoh et al., 2005). Multiple sequence alignment columns with a gap in more than 50% of sequences were deleted and not used in further analysis. Phylogenetic trees were calculated using FastME 1.1 (Desper and Gascuel, 2002) and RAxML 7.0.4 (Stamatakis, 2006), with the best-fit model inferred by ProtTest (Abascal et al., 2005). TreeDyn 198.3 (Chevenet et al., 2006) was used for tree visualization. All sequence, alignment, and phylogeny files are available upon request.

Cluster analysis of all retrieved TIR domains was based on pairwise sequence similarities using the clustering program CLANS (Frickey and Lupas, 2004) with a *P*-value cut off of 1e-7.

2.3. Multiple structural alignment

Partial Order Structure Alignment (POSA) (Ye and Godzik, 2005) was used to build the multiple structural alignment of the available TIR domain structures, including bacteria TIR (from *Paracoccus denitrificans*, PDB 3H16 (Chan et al., 2009)), plant TIR (from *Arabidopsis thaliana*, PDB 3JRN (Chan et al., 2010)), and animal TIRs (from human: TLR1, PDB 1FYV (Xu et al., 2000); TLR2, PDB 1FYW (Xu et al., 2000); TLR10, PDB 2J67 (Nyman et al., 2008); IL-1RAPL, PDB 1T3G (Khan et al., 2004); MyD88, PDB 2JS7). The superimposed structures were displayed with Discovery Studio (http://accelrys.com).

3. Results

3.1. Animal SARM is closely related to bacterial TIR domain-containing proteins

Following a protocol described in Section 2, we identified a total of 1688 TIR domain-containing proteins, including 483 from bacteria, 11 from archaea, 1193 from eukaryotes, and 1 from viruses (a bacteriophage). The TIR domains from these proteins were aligned and used to calculate their evolutionary history, as presented in Fig. 1A. The evolutionary tree consists of three main branches, which mostly correspond to the phylogenetic division between the animal (cyan), the plant (green), and the bacteria (magenta) branches (see the discussion below for the exceptions to this rule). The topology of the tree suggests a very ancient origin of this family, followed by massive lineage-specific expansions in each branch. Internal structure of the plant and the animal branches of the tree shows early emergence of few major lineages of TIR domains, followed mostly by speciation events. In contrast, the bacterial branch of the tree has clearly evolved by multiple lateral transfer events (see Fig. 2 and the discussion below for more details).

The tree presented in Fig. 1A is built on a multiple alignment of a very divergent protein family. For this reason, the statistical significance of some specific branches is not high, and trees with slightly different topologies can also be built. However, the main observations discussed here are supported by all alternative tree topologies we investigated. To further verify these observations, we used the protein clustering program CLANS (Frickey and Lupas, 2004), which exploits all-against-all similarity, so it does not depend on the approximations inherent in building the phylogenetic tree. CLANS presents the internal structure of a protein family as a two dimensional graph. While CLANS result cannot be interpreted in the context of phylogeny, it can support (as in this case) the overall features of the phylogenetic tree. The plant and animal TIR domains form very tight clusters (green and cyan clusters on the top and bottom of Fig. 1B, respectively).

On closer inspection it can be seen that several eukaryotic TIR domain subfamilies are found within the bacterial branch of the tree (Fig. 2). Most notable is the presence of the TIR domains from the family of animal SARM innate immunity adaptors deep in the bacterial branch of the tree. This is also seen in the clustering results (Fig. 1B), where the SARM family forms a small patch (center left in Fig. 1B), clearly separate from the main cluster of animal TIR domains (bottom right in Fig. 1B). This suggests that the evolutionary history of the animal SARM family is different from that of other adaptors, with this family likely representing a lateral gene transfer between bacteria and animals. A multiple sequence alignment of selected TIR domains from animal SARM adaptors and bacterial TIR-containing proteins that are close in the TIR phylogeny (Fig. 1A and Fig. 2) are presented in Supplementary Fig. 1 to illustrate the similarity between them at amino acid level. In humans, SARM is the only TIR adaptor that negatively regulates TLR signaling, blocking TRIF-dependent NF-κB and interferon-regulatory factor 3 (IRF3) activation (Carty et al., 2006). It can also inhibit TRIF- and MyD88-mediated activator protein 1 (AP-1) activation and p38 phosphorylation (Peng et al., 2010). Amphioxus SARM—bbtSARM—can attenuate the TLR signaling via interaction with amphioxus MyD88 and tumor necrosis receptor associated factor 6 (TRAF6) (Yuan et al., 2010). CrSARM, the ortholog of human SARM from horseshoe crab, can inhibit the TLR signaling pathway via TRIF, suggesting the conservation of the negative regulatory function of SARM from arthropod to human (Belinda et al., 2008). It is intriguing to speculate that the negative regulatory mechanism of SARM may reflect evolutionary conserved function of bacterial TIR-containing proteins of inhibiting their hosts' immune response.

3.2. SARM is not alone—other animal TIR-containing proteins in bacterial branch

Besides the SARM family, several other adaptor-like proteins from invertebrate- and choanoflagellate-specific families are found on the bacterial TIR branch of the phylogenetic tree (Fig. 2), including: reversed SARM-like proteins from sea urchin, sea anemone,