STATE-OF-THE-ART REVIEW

Death domain fold proteins in immune signaling and transcriptional regulation

Yu-San Huoh\textsuperscript{1,2} and Sun Hur\textsuperscript{1,2}

\textsuperscript{1} Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute at Harvard Medical School, Boston, MA, USA
\textsuperscript{2} Program in Cellular and Molecular Medicine, Boston Children’s Hospital, MA, USA

Death domain fold (DDF) superfamily comprises of the death domain (DD), death effector domain (DED), caspase activation recruitment domain (CARD), and pyrin domain (PYD). By utilizing a conserved mode of interaction involving six distinct surfaces, a DDF serves as a building block that can densely pack into homomultimers or filaments. Studies of immune signaling components have revealed that DDF-mediated filament formation plays a central role in mediating signal transduction and amplification. The unique ability of DDFs to self-oligomerize upon external signals and induce oligomerization of partner molecules underlies key processes in many innate immune signaling pathways, as exemplified by RIG-I-like receptor signalosome and inflammasome assembly. Recent studies showed that DDFs are not only limited to immune signaling pathways, but also are involved with transcriptional regulation and other biological processes. Considering that DDF annotation still remains a challenge, the current list of DDFs and their functions may represent just the tip of the iceberg within the full spectrum of DDF biology. In this review, we discuss recent advances in our understanding of DDF functions, structures, and assembly architectures with a focus on CARD- and PYD-containing proteins. We also discuss areas of future research and the potential relationship of DDFs with biomolecular condensates formed by liquid–liquid phase separation (LLPS).

Introduction

Numerous recent examples have shown that the formation of aggregate-like biomolecular assemblies is necessary for the facilitation of various cellular processes [1–3]. Such assemblies include macromolecular structures formed during many signal transduction pathways and membrane-less granules formed by

Abbreviations

AIM2, absent in melanoma 2; Aire, autoimmune regulator; Apaf-1, apoptotic protease-activating factor 1; ASC, apoptosis-associated speck-like protein containing a CARD; BCL10, B-cell lymphoma/leukemia 10; BCR, B-cell receptor; CARD, caspase activation recruitment domain; CARMA1, caspase recruitment domain-containing membrane-associated guanylate kinase protein-1; DD, death domain; DDF, death domain fold; DED, death effector domain; HIN, hematopoietic expression, interferon-inducible nature, and nuclear localization domain; IDR, intrinsically disordered region; IFI16, interferon-γ-inducible protein 16; IRAK, interleukin-1 receptor-associated kinase; LLPS, liquid–liquid phase separation; MALT1, mucosa-associated lymphoid tissue 1; MAVS, mitochondrial antiviral signaling protein; MDA5, melanoma differentiation-associated protein 5; MNDA, myeloid cell nuclear differentiation antigen; MyD88, myeloid differentiation primary response 88; PML, promyelocytic leukemia protein; PYD, pyrin domain; PYHIN, pyrin and HIN domain; RIG1, retinoic acid-inducible gene 1; RIP2, receptor-interacting serine/threonine-protein kinase 2; RLR, RIG-I-like receptor; Sp100, speckled protein 100; TCR, T-cell receptor; TRAF, tumor necrosis factor receptor-associated factor.
molecular phase separation. These are aggregate-like because they form large supramolecular assemblies (> 100-mer), and their turnover often requires cellular mechanisms used for misfolded protein aggregates [4,5]. However, these assemblies differ from misfolded protein aggregation in that they are mediated by properly folded proteins and/or nucleic acids and form specifically in response to cellular stimuli. These aggregate-like assemblies are thought to facilitate biological processes through local concentration of enzymes or signaling molecules for more efficient activities or sequestration of biomolecules from the rest of the cellular environment [1,3].

A common protein motif that mediates the formation of these functional macromolecular assemblies is the death domain fold (DDF) superfamily. The DDF superfamily comprises of 4 members: death domain (DD), death effector domain (DED), caspase activation recruitment domain (CARD) and pyrin (PYD) [6]. DDs and CARDs have been identified in apoptotic proteins of C. elegans and drosophila and likely emerged to mediate primordial cell death pathways [7]. DEDs and PYDs potentially diverged from these early apoptotic proteins as DEDs/PYDs have only been identified in vertebrates to mediate immune processes [6,7]. As DDFs have been mainly characterized in cytosolic innate immune and cell death pathways, it is entirely possible that there are still cryptic DDFs that govern other cellular processes yet to be identified.

All DDF members share similar fold and mediate homotypic protein:protein interactions resulting in the formation of high-order oligomers [8]. In isolation, these high-order oligomers can organize into repeating elements with helical symmetry that can be visualized as filaments [8]. Filamentous DDF assemblies have been well studied in the context of various vertebrate immune processes including foreign nucleic acid sensing, inflammation, cell death, and immune cell activation [2,9,10]. More recently, DDF proteins have been shown to also mediate transcriptional regulation of immune processes [11–13]. However, the functions of DDF multimerization within the nucleus are only beginning to be realized. As deeper mechanistic studies of nuclear DDFs are warranted, we discuss seminal studies used to dissect the mechanisms of cytoplasmic DDFs in hopes to apply the same approaches to characterizing structures and functions of nuclear DDFs.

In this review, we will focus on the studies of two subclasses of the DDF superfamily, the CARDs and PYDs, which broadly cover many of the diverse functions of all DDF members. We will also discuss higher-order assemblies of DDFs juxtaposed to liquid–liquid phase separation (LLPS) of biomolecules, a concept recently adopted to describe the organization of an increasing number of molecular condensates and aggregate-like assemblies.

The structure and high-order assemblies of DDFs

Biophysical and biochemical properties of DDFs

Death domain folds are characterized by common fold of six α-helices arranged in a Greek key conformation [8] (Fig. 1A). Structural studies have shown that α-helical bundling of DDF is governed by the packing of conserved hydrophobic residues from all six α-helices, whereas variations within the α-helices (termed H1-H6) define the subfamily members [8]. Compared with other DDFs, CARDs have an extended H1, which can sometimes contain a small kink or break that divides this helix into H1a and H1b (Fig. 1A). PYDs, on the other hand, have the shortest H3 among DDFs preceded by a flexible H2-H3 loop of varying lengths (Fig. 1A). With only a few noted exceptions (e.g., MALT1 DD interacting with BCL10 CARD [14]), these helical differences between DDFs appear to be sufficient to confine homotypic DDF interactions to their respective subfamily members [8].

Several challenges have previously hampered structural and biochemical studies of DDFs. Despite similar helical bundling and topologies, DDFs share low sequence homology even among subfamily members [6] (Fig. 1B). In fact, secondary or tertiary structure predictions are often required to identify these domains [11,15]. It is very likely that many new DDFs remain to be uncovered. Along with the difficulty of identifying DDFs, DDF-containing proteins display a high aggregation propensity and have been considered challenging for biochemical and structural characterization. Early biochemical studies mistook large aggregates as improperly folded and thus treated these assemblies as artifacts. To better handle these aggregating proteins, extreme nonphysiological conditions (e.g., low pH) were frequently used to maintain the monomeric state of DDFs [16–18]. However, subsequent structural studies have now cautioned the use of nonphysiological conditions and protein engineering as these can alter the conformation of DDF and induce formation of artificial oligomeric structures [19–21].

Filament assemblies of DDFs

Multiple studies of isolated DDFs demonstrate that the apparent aggregation property stems from the intrinsic propensity to self-associate into highly
Fig. 1. The structures and filamentous assemblies of DDF proteins. (A) DDF secondary structure topology diagram of six α-helices (H1-6) in a Greek key conformation. (B) Sequence alignment of CARD domains from human proteins listed. The alignment was generated with PROMALS3D and Esprit 3. (C) Negative stain EM image of MAVS<sup>CARD</sup> filament. (D) Surface and ribbon representations of MAVS CARD filament structure. (E) Structural model and cartoon representation of the DD-mediated Myddosome. The MyD88<sup>DD</sup> forms a hexamer, which serves as a template for IRAK4<sup>DD</sup> oligomerization. Downstream kinases such as IRAK2 can then engage with the MyD88:IRAK4 complex also through DD interactions. (F) Structural model and cartoon representation of the CARD-mediated RIG-I signalosome. The RIG-I 2CARD forms a tetramer, which serves as a template for MAVS<sup>CARD</sup> filament formation. The model was generated by superimposition of PDB IDs 3J6J and 4P4H. Figure 1C was reproduced from Ref. [46].
ordered helical assemblies (Fig. 1C). The first observation that DDFs form higher-order assemblies of potential infinite length was the Myddosome, consisting of DDs of MyD88 and IRAK kinases assembled into helical multimers [22]. Since then, many high-resolution structures of DDF oligomers and filaments have been reported [8] (Fig. 1D). These structures commonly show that DDFs form helical homomultimers (or filaments) using 3 conserved paired modes of binding involving 6 interfaces (referred to as type I-III interactions) [8]. The type I-III interactions are predominantly mediated by electrostatic or hydrophobic interactions of surface residues. Type I is mediated by residues on H1/H4 of a DDF with residues on H2/H3 of the cognate interaction partner; type II consists of the H4/H5 loop interacting with the partnering H5/H6 loop; and type III involves the H1/H2 loop interacting with the partnering H2/H3 loop. Despite the conserved set of interactions between DDFs, precise helical assemblies of DDFs vary depending on the shape and charge distribution of DDFs.

One of the unique characteristics of DDFs is that they not only form homomultimers, but also these homomultimers can induce the multimerization of DDFs within their interaction partners (Fig. 1E, F). This was again first demonstrated with Myddosome assembly (Fig. 1E) [22]. When MyD88 DD is left in isolation, it forms self-propagating homofilaments [23]. However, in the presence of IRAK4 and IRAK2 DDs, six to eight MyD88 DDs form a helical oligomer. This MyD88 helical oligomer recruits four IRAK4 DDs and four IRAK2 DDs to form the next layers of the helical assembly. Since upstream TLRs induce stable oligomerization of MyD88, this helical assembly mechanism allows a signal transduction from TLRs to MyD88, and to IRAK4 and IRAK2 [22,23]. A similar mechanism has been seen with RIG-I-like receptors (RLRs; Fig. 1F), inflammasome, and other immune signaling pathways (see below), placing DDF assemblies at the center of immune signal transduction. This mechanism also allows directionality and specificity of signaling, which will be discussed in detail later.

**Functions and regulation of DDF filament formation**

Structure-based mutational analyses have been crucial for elucidating the functions of DDF filament formation. There are now many examples linking filament formation to protein function *ex vivo* and *in vivo*. Mutations that disrupt DDF filament formation *in vitro* have been shown to also cause the loss of ASC-mediated inflammasome formation, downregulation of RIP2-mediated inflammatory signaling, MAVS-stimulated antiviral signaling, and Aire-mediated transcriptional activity [12,24–27]. Corroborating with these experiments, mutations that disrupt DDF multimerization provide the molecular mechanisms for certain human diseases; for example, autoimmune polyendocrinopathy syndrome type 1 (APS-1) can be caused by mutations in Aire CARD that disrupt Aire polymerization [12].

While isolated DDF can spontaneously form filamentous aggregates, many proteins with DDFs are spatially and temporally regulated to ensure multimerization occurs only at the right time and place [2,3]. To prevent spontaneous multimerization of DDFs, some proteins are in an autoinhibited state either by post-translational modifications that prevent DDF:DDF interaction or by a regulatory domain or binding partner that masks the DDF:DDF interface. In many cases, however, specific autoinhibition mechanisms are unknown, and maintaining the protein at low basal concentration is thought to be sufficient to prevent spontaneous multimerization.

**DDF filament vs. foci formation**

If filament formation is important for cellular functions, can one visualize the DDF filaments in cells? This has turned out to be challenging to address. While certain DDFs in isolation show filament-like fibril structures upon overexpression [28–30], endogenous DDF-containing proteins instead display round foci or punctate structures. Additionally, in some cases, fusing DDF protein with a bulky fluorescent tag altered the shape of filaments and other properties of the foci [12,19,31], demanding caution and carefully controlled experiments for visualization of DDF assemblies within cells. That being said, multiple immunofluorescence microscopy studies suggest that puncta formation of DDF-containing proteins is likely mediated by filament formation. These include MAVS [27], RIP2 [25] and ASC [24,26] showing cytoplasmic foci, and Aire showing nuclear foci in medullar thymic epithelial cells (mTECs) [32]. In these examples, mutations that disrupt *in vitro* filament formation generally led to loss of foci formation and respective functions in cells. Super-resolution microscopy further links DDF filament formation *in vitro* with foci formation in cells. External signals such as *Salmonella* infection can lead to the formation of the inflammasome, visualized as an ASC ‘speck’, within macrophages [33]; this *Salmonella*-induced ASC speck contains the DDF proteins NLRC4, NLRP3, ASC, and caspases 1 and 8—all of which have been shown to form filaments...
in vitro [29,34,35]. In another instance of inflamma-
some visualization, densely packed ASC could be iden-
tified to be concentrated at the nuclear membrane
periphery using electron microscopy of ultra-thin cryo-
COS-1 cell cross sections [30]. Collectively, these stud-
ies support the notion that DDF filament formation
drives their foci formation in cells and that these foci
are the sites of function, rather than of sequestration
or passive storage. The lack of visible linearity or fib-
rils in DDF puncta is presumably the result of the
dense packing or tangling of multiple DDF filaments
along with other protein interaction partners recruited
to these sites.

**DDFs in immune signaling**

Now that we have introduced basic concepts of DDF
structure and assembly, we describe in detail a few
examples with a focus on CARD- and PYD-
containing proteins and the functions of their filament
formation (Figs 2 and 3).

**CARDs in innate immune pattern recognition—
RIG-I-like receptors**

Death domain folds are common in various pattern
recognition receptors and innate immune signaling
pathways [6]. One of the earlier examples where the
functions and mechanisms of CARD are best charac-
terized is the RIG-I-like receptor (RLR) pathway
(Fig. 3A). RIG-I and MDA5, two members of the
RLR family, are cytoplasmic sensors of foreign
dsRNAs [36]. RLR sensing of foreign RNA can lead
to the activation of an antiviral signaling pathway that
culminates in the production of type I interferons
(IFNs) [36]. RIG-I and MDA5 contain tandem
CARDs (2CARD) for signaling and a helicase domain
followed by a C-terminal domain (CTD) both for
RNA recognition [37]. In the absence of RNA ligands,
RIG-I CARD is autoinhibited by an intramolecular
interaction with the helicase-CTD [38]; for MDA5, it
is unclear whether an analogous autoinhibitory state
exists, but monomeric MDA5 is inactive presumably
by maintaining a low basal concentration of the pro-
tein. Unless MDA5 and RIG-I are overexpressed, the
presence of uninhibited RLR CARDs is insufficient to
activate downstream signaling pathways. In order to
signal, RIG-I and MDA5 at physiological concentra-
tions require RNA-mediated homo-oligomerization of
the RNA-binding domain (helicase-CTD) along their
respective dsRNA ligands. This RNA-binding activity
promotes homotetramerization of RLR CARDs in
two synergistic ways. First, multimeric assembly of
RIG-I or MDA5 helicase-CTD along dsRNA places
the tandem CARDs in close proximity, which pro-
motes CARD tetramer formation [39,40]. In addition,
filament formation of MDA5 and RIG-I helicase-
CTD:dsRNA enables avidity-dependent recruitment of
E3 ubiquitin ligases that conjugate K63-linked polyu-
biquitin, TRIM65 [41] and RIPLET [42], respectively.
Structural studies showed that K63-linked polyubiqui-
tin chains bind the tandem CARD tetramer of RIG-I
[43], which stabilizes the CARD tetramer required for
consequent signaling activity. While unanchored K63-
linked ubiquitin chains can enhance CARD tetramer-
ization in vitro, covalent anchoring is likely required in
cells as it greatly facilitates CARD:ubiquitin binding
and ubiquitin-mediated CARD tetramerization [42,43].

![Fig. 2.](image-url) Domain architecture of the DDF proteins featured in this review. The death domain fold (DDF) is a common interaction motif that
can be found in combination with catalytic domains (e.g., kinase, caspase) and other interaction motifs (e.g., Bromo, SH3). This review
features both cytosolic and nuclear DDF proteins that mediate a variety of biological processes including immune signaling, cell death
pathways, and transcriptional regulation.
How does the tandem CARD tetramer of RIG-I/MDA5 activate downstream signaling in a multimerization-dependent manner? RIG-I and MDA5 share a common signaling adaptor molecule, MAVS, which contains an N-terminal CARD and a C-terminal transmembrane (TM) domain that anchors to the outer mitochondrial membrane [44, 45]. Sequence homology analysis suggests that MAVS CARD is most closely related to CARDs of RIG-I and MDA5 [46]. Structural studies further explain why the RIG-I and MDA5 2CARD tetramers are uniquely capable of recruiting MAVS CARD and inducing MAVS CARD filament formation [43, 46]. RLR 2CARD tetramer has a helical symmetry that recruits MAVS CARD along this predefined helical trajectory, thereby serving as a template for MAVS filament formation [43, 46]. With this mechanism, a small number of activated RLRs can induce oligomerization of many MAVS molecules, allowing rapid amplification of the signal. Although MAVS CARD can form filaments of infinite length in vitro, super-resolution microscopy studies suggest that the size of MAVS aggregates is limited to ~150 MAVS molecules per fibril [47]. Nevertheless, MAVS filament of this size is sufficient for activating downstream signaling pathway [47] by serving as a signaling platform to recruit and activate TRAF2/3/5/6, TBK1, and other signaling molecules [48, 49].

One of the intriguing aspects of RLR-MAVS CARD:CARD interaction is its directionality and specificity that confines these characteristics to the RLR signaling pathway. First, while RLR CARD tetramers induce MAVS filament formation, the reverse does not appear to occur. That is, there is no evidence suggesting that activation of RIG-I leads to activation of MDA5.
through MAVS or vice versa. The CARD:CARD interaction between RLRs and MAVS is also highly specific [46]. RLR activation does not generally lead to the activation of other CARD- or DDF-containing proteins, although some reports suggest such crosstalk [50–53]. Despite having MAVS CARD as a common interaction partner, RIG-I CARD and MDA5 CARD also do not interact with each other. Together with amplificability, directionality and specificity of homotypic interactions are likely common properties shared among many DDFs that make DDFs ideal building blocks for signal transduction.

**CARDs/PYDs in innate immune pattern recognition—Inflammasomes**

Inflammasomes are large signaling complexes consisting of the receptor or receptor-like molecule (such as NLRP1 or AIM2), the adaptor ASC, and the effector inflammatory caspase (such as caspase-1) [54]. These complexes assemble in the cytoplasm in response to a variety of self- and non-self-derived insults and stimuli [54]. Inflammasome assembly activates the proteolytic activity of inflammatory caspases, which subsequently leads to the processing of the cytokines IL-1β and IL-18 along with pore-forming gasdermin D that triggers pyroptosis [54]. In many cases, the mechanisms of how certain stimuli activate specific inflammasome assemblies remain unclear.

One of the better characterized ‘canonical’ inflammasome assembly mechanisms is mediated by the foreign dsDNA sensors AIM2 and IFI16 [2] (Fig. 3B). AIM2 and IFI16 belong to the PYHIN family, a group of proteins that contain an N-terminal PYD and at least one HIN domain, a sequence-independent DNA binding module [55]. AIM2 surveys the cytoplasm where host DNA is normally absent [56–58] and thus likely discriminates self vs. nonself DNA on the basis of self DNA exclusion from the cytoplasm. By contrast, IFI16 has been found in both the cytoplasm and nucleus [59] and relies on the recognition of long stretches of nonchromatinized DNA to discriminate nonself from self DNA [60]. For AIM2 and IFI16, there does not appear to be an autoinhibition mechanism analogous to that of RIG-I to suppress spontaneous oligomerization of PYD and these receptors may rely on low basal protein concentrations to maintain the inactive state [31,61]. Once respective foreign DNA is detected by multiple AIM2/IFI16 HINs, the dsDNA:HIN complexes presumably increase the propensity of their PYD to multimerize [31,60,61]. Reciprocally, AIM2/IFI16 DNA binding by HINs is further enhanced by AIM2/IFI16 PYD multimerization [31,61]. IFI16/AIM2 PYD multimerization then initiates a cascade of DDF homo- and hetero-oligomerization that leads to inflammasome assembly. First, the AIM2/IFI16 PYD multimer serves as the helical template for recruitment and filament formation of ASC PYD (this mechanism mirrors how RIG-I induces MAVS filament formation) [31,60,61]. Second, ASC PYD multimerization in turn enables proximity-induced oligomerization of the C-terminal CARD of ASC [24,30]. Third, the CARD multimer of ASC then nucleates the CARD on pro-caspase-1, promoting proximity-based activation of caspase-1 [19,29]. Finally, a proper inflammatory response results after activated caspase-1 proteolytically processes pro-IL-1β and pro-IL-18 for NF-κB signaling activation and also cleaves gasdermin D for pyroptosis initiation [54]. Once formed, the inflammasome complex comprising of receptor:ASC:pro-caspase-1 can be visualized as a single macromolecular aggregate known as the ASC speck in cells [54]. Thus, for both RLR signalosome and inflammasome, a chain of DDF filament formation mediates signal activation and transduction that can be visualized as foci in the cell.

**DDFs in lymphocyte immune signaling**

In addition to being critical for mediating innate immune signaling, DDF filament formation is crucial for mediating lymphocyte immune signaling. For instance, T-cell and B-cell antigen receptor (TCR/BCR)-mediated NF-κB signaling involves DDF filament formation that shares regulatory mechanisms common to both RLR signalosome and inflammasome assembly [9] (Fig. 3C). Following antigen peptide: MHC engagement and TCR/BCR activation, a succession of phosphorylation events occurs, one of which leads to the hyperphosphorylation of a flexible linker within CARD membrane-associated guanylate kinase protein 1 (CARMA1, also known as CARD11) [62,63]. In the resting state of T and B cells, CARMA1 CARD is inhibited by intramolecular interactions mainly through its linker region [63]. Hyperphosphorylated CARMA1 linker undergoes a conformational change, which releases autoinhibition of an N-terminal CARD and allows for its oligomerization [62,63]. Oligomerized CARMA1 CARD then functions as a nucleator for the filament formation of the CARD on the adaptor molecule B-cell lymphoma 10 (BCL10) [64]. As BCL10 is constitutively bound to the paracaspase MALT1 in a 1:1 stoichiometry, BCL10 filament formation brings MALT1 in proximity to another MALT1 molecule, thereby stimulating MALT1 paracaspase activity [14]. The proteolytically active
CARMA1/BCL10/MALT1 supramolecular assembly is known as the CBM complex and is the signaling platform that recruits downstream E3 ubiquitin ligases and kinases responsible for mediating NF-κB downstream signaling events within lymphocytes [65]. Much like for RLR signalosome and inflammasomes, CBM complex formation is a tightly controlled process involving a substoichiometric nucleator, which, once oligomerized, can respond to stimuli and amplify signals rapidly through the recruitment of partners into helical filamentous assemblies.

**CARDs/PYDs in gene regulation**

While cytoplasmic DDF assemblies have been extensively studied in immune signaling, functions of nuclear DDFs in transcriptional regulation are only beginning to emerge. In this section, we discuss the currently known functions of CARD/PYD transcriptional regulators and pose questions that could reveal potential molecular mechanisms behind these more elusive nuclear filamentous assemblies (Figs 2 and 4).

**Speckled protein 100 (Sp100) in transcriptional regulation and antiviral immunity**

The Sp100 family proteins (Sp100, Sp110, Sp140, Sp140L, and Aire) are nuclear proteins that contain CARD domains and form nuclear foci [66]. These proteins also have adjacent chromatin interaction and DNA binding motifs, suggesting that they play important roles in regulating gene expression [66]. With the exception of Aire, functions of all other Sp100 family members are poorly characterized. Sp100, Sp110, and Sp140 have been shown to colocalize with nuclear molecular condensates known as promyelocytic leukemia bodies (PML bodies, also known as ND10 bodies) [67–69]. This suggests that the PML-associated Sp100 family members may have functions linked to those of PML. PML is an E3 SUMO tripartite motif (TRIM) ligase family member and interacts with a variety of substrates and binding partners in order to form PML bodies [70]. PML bodies have been reported to mediate a diverse set of cellular functions including nuclear protein quality control, DNA damage, apoptosis, cell cycle, stress responses, and innate immunity [70], although detailed mechanisms for many of these proposed functions are unclear. Sp100, Sp110, Sp140, and Sp140L have been shown to be transcriptionally upregulated with type I IFN [68,69,71,72], raising the possibility that they may function in antiviral defense.

This seems to be particularly the case for Sp100, which has been reported to restrict in cell lines a variety of viruses including human papillomavirus (HPV), herpes simplex virus-1 (HSV-1), cytomegalovirus (CMV), adenoviruses, and retroviruses [73–77]. The ability of Sp100 to regulate a wide variety of viruses suggests that Sp100 inhibits viruses by intrinsic and/or innate mechanisms. It has been postulated that a general antiviral function of Sp100 could be to maintain the structural integrity of PML bodies by continuously being SUMOylated and bound to PML and other PML-associated factors [76,78]. By maintaining the structural integrity of PML bodies, Sp100 may concentrate and retain the epigenetic restriction factors that immediately encounter viral components [76,79]. The potential role of CARD multimerization in PML association and antiviral immunity remains unclear. In addition, some Sp100 family members share high CARD sequence homology [72], it is tempting to speculate that Sp100 family members interact with each other or with common partners through CARD:CARD interactions in a manner similar to those observed for the RLR signalosome and inflammasome assemblies.

**Aire multimerization in the transcriptional regulation of T-cell tolerance**

Aire is a transcriptional regulator that plays a key role in establishing central tolerance in T-cell immunity [80]. Aire does so by being expressed in thymus medullary epithelial cells and upregulating thousands of tissue-specific antigens (TSAs), which are presented to developing thymocytes for negative T-cell selection [81]. Accordingly, mutations in AIRE cause the multi-organ autoimmune disease APS-1 [82]. While the biological function of Aire is well characterized, the mechanism by which Aire regulates TSA expression is still unclear. Studies showed that Aire lacks the ability to directly bind DNA, but instead appears to recognize and alter certain chromatin structures, such as superenhancers [83], thereby indirectly controlling expression of a large number of target genes.

Aire harbors a CARD domain at the N terminus, which was annotated based on its sequence similarity to Apaf-1 CARD [11]. Aire CARD is frequently mutated in APS-1 patients and was shown to play an important role in Aire’s transcriptional activity [11]. Mutation analyses suggest that Aire foci formation is important for Aire’s transcriptional function [11]. Unlike other members of the Sp100 family, Aire foci do not colocalize with PML bodies or other previously known subnuclear structures [84]. A more recent study showed that Aire CARD forms filament in vitro and this property can explain Aire foci formation [12] (Fig. 4A,B). Disruption of Aire filament formation...
correlates with loss of transcriptional activity and diffuse nuclear staining.

Intriguingly, Aire CARD can be functionally substituted with tandem repeats of a chemically inducible oligomerization domain, FKBP, suggesting that filament architecture is unnecessary for Aire’s function [12]. This is consistent with the notion that the role of Aire CARD is to promote the functions of other parts of Aire, such as the PHD and activation domain (AD), which, by an avidity effect, can upregulate the expression of tissue-specific antigens. However, not all Aire aggregation is the same; in fact, Aire CARD-mediated multimerization also makes Aire susceptible to interaction with promyelocytic leukemia protein (PML) bodies, sites of many nuclear processes including protein quality control of nuclear aggregates. Aire aggregates localized at PML bodies are dysfunctional and can exert a dominant negative effect by recruiting other Aire molecules to PML bodies. Fig. 4A,B were reproduced from Ref. [12].

How is Aire CARD multimerization controlled and what happens when it occurs in an inappropriate manner? While the precise regulatory mechanism is still unclear, a recent study suggests that actions of the nuclear protein quality control system suppress transcriptional activity of improperly multimerized Aire.

Fig. 4. Aire CARD multimerization is critical for transcriptional regulation of T-cell tolerance. (A) Immunofluorescence microscopy image of Aire (green foci) in 4D6 cells (epithelial thymic cell line). Nuclei were visualized with DAPI stain (blue). (B) Representative negative stain EM image of Aire CARD filaments. While fully formed CARD filaments (i) were major species, thin protofilaments (ii) were also observed. Right, representative images of the two types of filaments. 2D class averages were also shown for fully formed CARD filaments. (C) Aire CARD multimerization is spatially and temporally regulated. Upon entry into the nucleus, Aire can receive a stimulus to promote its filament formation. Properly formed Aire filaments likely promote the functions of other parts of Aire, such as the PHD and activation domain (AD), which, by an avidity effect, can upregulate the expression of tissue-specific antigens. However, not all Aire aggregation is the same; in fact, Aire CARD-mediated multimerization also makes Aire susceptible to interaction with promyelocytic leukemia protein (PML) bodies, sites of many nuclear processes including protein quality control of nuclear aggregates. Aire aggregates localized at PML bodies are dysfunctional and can exert a dominant negative effect by recruiting other Aire molecules to PML bodies. Fig. 4A,B were reproduced from Ref. [12].
Several loss-of-function Aire mutations, including those causing APS-1, misdirect Aire to PML bodies, sites of protein quality control (PQC) of nuclear aggregates, and consequently impair Aire transcriptional activity. Targeting Aire to PML bodies inhibits Aire’s transcriptional activity. Conversely, dispersing PML bodies with a viral antagonist can partially restore transcriptional activity for Aire variants that associate with PML bodies. Furthermore, when coexpressed with wild-type Aire, PML-localizing mutants redirect wild-type Aire to PML bodies, explaining their dominant negative activity in APS-1 patients. Finally, by virtue of forming large homomultimeric, aggregate-like assemblies, even wild-type Aire seems subject to PML-mediated surveillance, although not stably localizing at PML bodies. Thus, Aire CARD appears to function as a double-edged sword that confers both the critical function in Aire’s activity and susceptibility to PQC surveillance (Fig. 4C). It also suggests that Aire normally employs as-yet-unknown mechanisms to avoid stable PML localization and transcriptional suppression.

Clues about the precise regulatory mechanism of Aire CARD polymerization may be obtained from more detailed structural analysis of Aire and Aire CARD filaments. While there are numerous high-resolution filament structures of DDFs in immune signaling pathways, thus far, there are no structures of nuclear DDF filaments. Negative stain EM analyses showed that Aire CARD filaments display features distinct from those in the cytoplasmic signaling molecules [12]; Aire CARD filaments can exist in a ~15-nm-thick mature filament and a ~10-nm-thick protofilament (Fig. 4B). This differs from the MAVS CARD (Fig. 1C) and ASC PYD filaments that cooperatively form only mature filaments without apparent intermediate states [30,46]. Whether such stepwise filament assembly is common to other members of the Sp100 family and whether this allows for different functions and regulatory mechanisms that are distinct from cytosolic DDFs remain to be investigated. Furthermore, future high-resolution structures of Aire CARD and other nuclear DDF filaments would be necessary to understand how different filament architectures are assembled using the common DDF fold.

Other emerging functions of CARD/PYD-containing transcriptional regulators

Sp140 is emerging as a critical regulator of macrophage transcriptional programing [13]. Sp140 maintains macrophage identity by binding to repressed chromatin (H3K27me3 sites), thereby suppressing lineage-inappropriate genes (e.g., HOX genes). This transcriptional repression mechanism is in contrast to that of Aire, which involves association with superenhancers and activation of normally silenced genes that encode tissue-specific antigens. It is unclear how two related transcription factors with a similar domain architecture and a closely related CARD have such opposing functions. More studies are required to elucidate the functional role of CARD multimerization for Sp140-mediated transcriptional regulation.

Myeloid cell nuclear differentiation antigen (MNDA), another member of the PYHIN family, is the only PYD-containing nuclear protein currently known to be implicated in directly controlling transcription [86,87]. MNDA is a key lineage-specific transcriptional regulator particularly for the granulocyte/monocyte progenitor [88]. While the precise mechanism of how MNDA controls hematopoietic gene expression remains unknown, MNDA interaction with the ubiquitous transcriptional regulator Yin Yang 1 (YY1) may be relevant. MNDA:YY1 complex was shown to stimulate YY1 interaction with its target DNA [87]. This MNDA:YY1 interaction was narrowed down to the first 200 amino acids of MNDA, which includes the PYD [87]. Although NMR and X-ray crystal structures of MNDA PYD were determined in the monomeric state (PDB ID: 2DBG and [89], respectively), evidence suggests that MNDA PYD has the ability to multimerize like other PYDs. For example, MNDA PYD can functionally replace IFI16 PYD as the chimeric protein MNDA PYD-IFI16ΔPYD was demonstrated to form nuclear foci and possess antiviral activity like wild-type IFI16 [90]. Whether MNDA requires PYD multimerization to perform its transcriptional regulatory function remains to be determined.

DDF assemblies vs. biomolecular condensates

One of the common characteristics of DDF-containing proteins is their ability to form microscopically observable puncta or foci in the cell, a property shared with proteins that form biomolecular condensates by LLPS. In this final section, we juxtapose the properties and functions of biomolecular condensates with DDF assemblies in order to understand their potential relationships (Fig. 5).

DDF assemblies vs. biomolecular condensates in vitro

Biomolecular condensates with liquid-like behavior have been known for over 3 decades [91], but only...
recently have their potential functions been explored for a variety of proteins in diverse biological contexts [1]. As LLPS is an emerging concept in biology, the definition of biomolecular condensates has also continued to evolve and be refined. Nonetheless, the generally accepted definition of biomolecular condensates is that they are dense molecular structures (on the micron scale) without membrane-enclosed boundaries [1]. They often form when the valency of cognate interaction motifs, while weak, is enriched to the point where these interactions partition into two distinct phases, a dilute and dense phase where a biomolecule is enriched in one phase and excluded from the other [92,93]. As such, molecular condensates often display liquid-like properties, characterized by dynamic exchange with the surrounding environment and rapid dissolution of its structure upon dilution [94]. From this perspective, DDF assemblies differ from phase-separated condensates in that they, at least in vitro, are extremely stable to the point that their disassembly often requires chemical denaturation [48,95].

Another difference between DDF assemblies and condensates is specificity requirements. Structural studies have shown that DDF filament formation has a defined number of valencies and is highly sequence-specific [8]. Single amino acid substitutions that do not disrupt the structural integrity of DDFs can still disrupt filament formation [12,30]. By contrast, LLPS involves dynamic networks of weak interactions mediated by protein domains with low sequence complexity (e.g., acidic, proline-, serine/threonine- or glutamine-rich) or intrinsically disordered regions (IDRs) [1]. Thus, disruption of LLPS at times requires extensive mutagenesis; for instance, OCT4 condensate dissolution required 17 Asp/Glu-to-Ala substitutions within its IDR to see an effect on liquid droplet formation and transcriptional activity [96].

**DDF assemblies vs. biomolecular condensates in cells**

Despite the distinct biophysical properties of biomolecular condensates and DDF multimers, their differences are less clear in the context of complex cellular environment and much investigation is needed to define the relationship between DDF assembly and LLPS. For example, the dynamic properties of DDF assemblies in cells remain largely unknown due to the fact...
that microscopy experiments of filamentous foci have mostly been examined in fixed cells. It is possible that the DDF filament stability is dynamically regulated by various post-translational modifications, in much the same way that the stability of some biomolecular condensates has been proposed to be regulated [97] (Fig. 5). In addition, as with many multidomain proteins, DDF-containing proteins also often harbor loops or linkers that share common properties with IDRs [66,98]. Thus, in the context of full-length proteins, multimerization of DDF may result in linker/loop-mediated crosslinking between DDF filaments, resulting in the formation of higher-order structures or condensate-like entities. In fact, some DDF assemblies associate with known biomolecular condensates, either as core structural components or as passenger molecules. These include RLR signalosomes that are part of cytosolic stress granules [99,100], and the Sp100 family members that are part of PML bodies [67–69]. Precisely how DDF assemblies contribute to the formation of condensates, and vice versa, how LLPS affects the functions of the DDF assemblies are interesting topics for future investigation (Fig. 5).

Future directions, challenges, and concluding remarks

Led by the pioneering discovery of the Myddosome structure, many biochemical and structural studies of DDFs have now shown that DDFs generally form helical oligomers or filaments during their functions. By utilizing a conserved mode of interaction involving six distinct surface areas, DDFs function as a building block that can be densely packed to form homomultimers. By shielding one or more of the interfaces in the resting state or by simply maintaining low protein concentration, DDF proteins also regulate their multimerization so that it occurs only in response to cellular or environmental stimuli. Furthermore, using the same mode of interaction and interfaces, they can nucleate multimerization of another protein that shares a similar or structurally compatible DDF. This unique ability to self-oligomerize upon external signal and to induce oligomerization of a partner molecule underlies key processes in many innate immune signaling pathways, as exemplified by RLRs and inflammasome pathways. Such filament-mediated signal transduction and propagation allows rapid amplification of the cellular signal, which is likely essential for effective antimicrobial defense.

It remains to be investigated what role DDF plays in the functions of nuclear transcriptional regulators. Questions include how nuclear DDF multimerization occurs in a controlled manner and whether they also function by inducing multimerization of other DDF-containing partners? Additionally, we currently have limited understanding of how the DDF multimers, either in innate immune signaling or in transcriptional regulatory pathways, is resolved or degraded to restore cellular homeostasis, and, moreover, what role does LLPS play in either triggering DDF polymerization or clearing DDF polymers?

Another area of future research is the development of computational algorithms to identify new DDFs and predict their cognate partner DDFs. Based on the current structural and biochemical knowledge of the existing DDFs and their homotypic interaction network, it may be possible to uncover novel DDFs and DDF:DDF interactions in many other biological contexts besides immune signaling and transcriptional regulation. Such an effort would revolutionize our understanding of DDFs potentially beyond their already diverse roles as signaling molecules and assembly building blocks in immune signaling and transcriptional regulation.

Acknowledgements

The authors acknowledge NIH R01s (AI154653 and AI111784 to SH) and the Jeffrey Modell Foundation postdoctoral fellowship (YSH) for supporting this work.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

YSH and SH wrote the manuscript. YSH generated the included figures.

References

30 Lu A, Magupalli VG, Ruan J, Yin Q, Atianand MK, Vos MR, Schroder GF, Fitzgerald KA, Wu H &


