Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structural and Biochemical Characterization of the MAB21 Family Members and RIG-I Innate Immune Sensors

von

Carina Cristina de Oliveira Mann

aus

Lissabon, Portugal

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordung vom 28. November 2011 von Herrn Prof. Dr. Karl-Peter Hopfner betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 06. 05. 2016

.....

Carina Cristina de Oliveira Mann

Dissertation eingereicht am	06. 05. 2016
1. Gutachter	Prof. Dr. Karl-Peter Hopfner
2. Gutachter	Prof. Dr. Elena Conti
Mündliche Prüfung am	04. 07. 2016

This thesis has been prepared from February 2012 to April 2016 in the laboratory of Professor Dr. Karl-Peter Hopfner at the Gene Center of the Ludwig-Maximilians-Universität Munich.

This is a cumulative thesis based on following publications:

Civril, F.*, Deimling, T.*, de Oliveira Mann, C. C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.-P. (2013) Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332-337

de Oliveira Mann, C. C., Kiefersauer, R., Witte, G., and Hopfner, K.-P. (2016) Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Scientific Reports* **6**, 27498

Lässig, C., Matheisl, S.*, Sparrer, K. M.*, de Oliveira Mann, C. C.*, Moldt, M., Patel, J. R., Goldeck, M., Hartmann, G., García-Sastre, A., Hornung, V., Conzelmann K.-K., Beckmann R., Hopfner K.-P. (2015) ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *eLife* **4**, e10859

*: equal contribution

TABLE OF CONTENTS

ABSTRACT

Nucleic acids sensing pathogen recognition receptors (PRRs) are forced to develop sophisticated strategies to overcome self nucleic acid recognition. During evolution cytosolic nucleic acid sensors acquired elaborate mechanisms to recognize specific structural features confined to nucleic acids of pathogenic origin. Accurate understanding of these mechanisms and regulation of the innate immune response can be a key in therapeutic strategies for prevention and treatment of autoimmune diseases. In addition, information about how the innate immune response is activated and consequent modulation of the adaptive immune system is a valuable tool for the design of new immunostimulatory molecules for vaccination and for the constant fight against viral infections.

The cyclic GMP-AMP synthase (cGAS) was discovered as the major cytosolic sensor for dsDNA. Upon dsDNA recognition, cGAS catalyses the production of a novel second messenger cGAMP harboring a noncanonical 2´-5´ and a 3´-5´ mixed phosphodiester linkage. This work provided the crystal structures of porcine cGAS (MAB21 domain) alone and in complex with a 14-mer dsDNA and the substrates ATP and GTP together with functional assays *in vitro* and in cells. Our results revealed the cGAS activation mechanism explaining its broad sequence-independent sensing specificity along with an insight into the catalysis of the first linear intermediate. This structural work has an impact on our understanding of the evolutionary origin connecting cGAS to the dsRNA sensor oligo-adenylate synthetase 1 (OAS1). A new class of template-independent nucleotidyltransferases (NTases) acting as cytosolic nucleic acids sensing PRRs was revealed.

cGAS, previously known as MB21D1, is member of the NTase fold protein subfamily MAB21 which was named after a conserved family of eukaryotic cell fate determining proteins. MAB21-like proteins (MAB21L1, MAB21L2 and MAB21L3) have a fundamental role in development. Besides extensive characterization of drastic MAB21 phenotypes in different model organisms, little or nothing was known about the structure and molecular function of these proteins. The main aim of this work was to decipher the evolutionary path within members of the MAB21 family and cGAS in order to improve our enzymatic understanding of this newly discovered subfamily of NTases. This thesis provides a first structural framework for the highly conserved MAB21-like proteins. Crystal structures of full length human MAB21L1 alone and in complex with CTP were solved at 3.0 Å and 2.5 Å respectively. A high degree of structural conservation between MAB21L1 and cGAS was revealed along with some important differences. In the crystal structure MAB21L1 assembled into a highly symmetric double-pentameric oligomer, suggesting that oligomerization could be a feature of MAB21L1. Moreover, CTP did not bind in the NTase active site; instead CTP unveiled a potential ligand binding site. The drastic phenotypes of patient-related mutations in MAB21L2 and MAB21L1 could be explained by means of structural analysis, since affected residues were involved in interaction networks stabilizing the protein.

ABSTRACT

Finally, a part of this thesis dealt with structure-derived as well as patient-related ATPase mutants of retinoic acid inducible gene-I (RIG-I) and their specific effects *in vitro* and in human cells. Previous studies demonstrated that ATPase defective alleles of RIG-I and MDA5 resulted in autoimmune diseases including Aicardi-Goutières and Singleton-Merten syndromes. So far it was not clear how RIG-I like receptors (RLRs) make use of the driving force of ATP hydrolysis for translocation along the bound RNA. In this work, ATP hydrolysis was revealed to help RIG-I to discriminate between self and non-self dsRNA by fine-tuning the release of its tandem CARDs and the hereby associated downstream signaling events. Therefore low or impaired ATP hydrolysis in RIG-I and MDA5 mutants correlates with constitutive signaling inducing autoimmunity.

1 INTRODUCTION

1.1 THE INNATE IMMUNE SYSTEM: DISCRIMINATION OF SELF AND NON-SELF

Innate immunity represents the first line of defense that recognizes the invading pathogen and initiates host immune response through production of type I interferons and proinflammatory cytokines. The innate immune system was believed for a long time as being an evolutionary ancient remainder or fragment, whose function is only to rapidly isolate infection. Based on accumulating evidences, it is now a well-established paradigm that the innate immune system is essential for activation and subsequent guidance of adaptive immunity (2). It helps the adaptive immune response to discriminate between self and non-self by providing valuable information about the origin of the antigen, determining microbial presence, viability, replication, location and virulence. In addition, it ensures time for the clonal adaptive system to develop a specific and long-lasting immunity, while it controls the infection during the first critical period of 4-7 days (3).

In the course of evolution, this phylogenetically ancient defense mechanism present in all multicellular organisms, evolved a limited number of nonclonal, germ-line encoded receptors (4). Pathogen recognition receptors (PRRs) are constitutively expressed in the host and specialized in detecting one type of pathogen-associated (PAMP) or damage-associated molecular pattern (DAMP). PAMPs are molecular structures usually essential for the survival and/ or replication of the pathogen. These are shared by large groups of microorganisms and therefore invariant and difficult for the pathogen to mutate or exclude from detection without a fitness cost. Consequently, it is more complicated for the pathogen to avoid innate immunity and thus, most mechanisms to evade immune response are directed to the variable adaptive immune system. DAMPs in contrast, are originating from the host and can lead to inflammation when not found in the right context of a healthy cell, e.g. endogenous stress signals (5,6).

1.2 THE FIRST LINE OF DEFENSE: PATHOGEN RECOGNITION RECEPTORS

For the purpose of recognizing different microbial molecules, PRRs arose from a vast range of structures deriving from several protein families. PRRs sharing common protein domain structures, such as Toll-like-receptors (TLRs), rely on different amino acids compositions within these modules and on accessory molecules for diversity and specificity in ligand recognition (7). However, PRRs share a universal concept of initiating downstream signaling. After recognizing the respective ligand, PRR's signaling domains interact with downstream adaptor proteins resulting in similar signaling pathways activating the respective effector proteins. These lead to proinflammatory response followed by activation and recruitment of phagocytic cell as well as professional antigen-presenting cells. Principal functions of PRRs, besides activation of proinflammatory signaling, are therefore opsonization, activation of complement and coagulation cascades, phagocytosis and induction of apoptosis (8). Depending on the type of ligand they recognize, pathogen recognition receptors can be expressed on the cell surface, in intracellular compartments or secreted into the bloodstream and tissue fluids. There are several ways to classify PRRs: either by compartmentalization dividing into membrane-associated, cytoplasmic and secreted PRRs; tissue-specific expression or characterizing by structure and class of ligand recognized (9).

Mammalian Toll-like receptors (TLRs) were the first PRRs to be discovered and therefore the best studied (10). First evidences for this class of PRR came from the Toll protein in *Drosophila melanogaster*. Besides being already implied in embryonic dorso-ventral axis formation, Toll was shown to be responsible for effective immune responses in adult flies against the fungus *Aspergillus fumigates* (11,12). Following this discovery, a mammalian homolog of the Drosophila Toll receptor, namely Toll-like receptor 4 (TLR4), was identified. In the experiment a constitutively active mutant of human TLR4 was transfected into human cell lines and resulted in the activation of NK-κB, which led to the expression of genes encoding proinflammatory cytokines and upregulation of co-stimulatory molecules (10). The ultimate link between TLRs and the innate immune response was provided when a point mutation in the *TLR4* gene resulted in mice unresponsive to lipopolysaccharide (LPS) challenge and therefore more susceptible to Gram-negative bacterial sepsis (13). Hereby, the first study characterizing the ligand for a PRR demonstrated that TLR4 is the receptor for LPS. Subsequently, the generation of knockout mice for each member of the TLR family has revealed the respective PAMP recognized by each TLR.

So far 10 human TLRs and 13 in mice were identified. Each TLR recognizes a specific PAMP and this correlates with its subcellular localization. TLR recognizing molecular components located on the pathogen's surface reside at the plasma membrane (TLR1, TLR2, TLR4, TLR5, TLR6 and TLR11), while the ones recognizing nucleic acids are located in intracellular compartments, such as endosomes and lysosomes (TLR3, TLR7, TLR8 and TLR9) (14).

1.3 NUCLEIC ACIDS SENSING PATHOGEN RECOGNITION RECEPTORS

Nucleic acids, which are used as genetic information carrier by every organism, serve as PAMPs and expand the variety of pathogens recognized by the host innate immune system. The main difference in PRR recognition of nucleic acids as PAMPs is that they are not restricted to the pathogen. This blurs the lines between self and non-self discrimination and introduces the risk of host nucleic acid recognition by innate immune sensors which can lead to autoimmune and autoinflammatory diseases (1). Hence, nucleic acids sensing PRRs are forced to acquire new strategies to overcome self-recognition.

These PRRs are stringently compartmentalized in order to restrict accessibility to host nucleic acids and are responsible for surveillance of almost all cell types against invading pathogens (9). First characterized group of nucleic acid sensing PRRs were the Toll-like receptors (TLRs) that are expressed on the endosomal membrane where they monitor the lumen. Nucleic acids sensing TLRs are confined to cells of the immune system such as dentritic cells (DCs), macrophages and B cells. In contrast, non-immune cells such as epithelial cells and fibroblasts that reside on the mucosal surface rely on another group of PRRs. These cells, most prone to infection, express several types of cytosolic RNA and DNA sensors. In order to further avoid self-recognition, these cytosolic receptors developed several mechanisms to recognize defined biochemical features exclusively present on molecules of pathogenic origin. Nucleic acid receptors and their ligands will be described in detail below and focus will be kept on mechanisms and strategies of cytosolic nucleic acid sensing.

1.3.1 TOLL-LIKE RECEPTORS

TLRs are type I transmembrane proteins of the Interleukin-1 receptor (IL-1R) family. These integral membrane glycoproteins have a trimodular structure comprising an N-terminal ectodomain consisting of leucine-rich-repeats (LRR) for PAMP recognition, a single transmembrane domain and a C-terminal cytosolic Toll/ IL-1 (TIR) receptor domain for downstream signaling (15). The ligand binding domain of TLRs is confined to the lumen of endosomes, so that after endocytosis of the virus, the acidic environment is able to degrade viral particles followed by exposure of genomic nucleic acids and detection by the respective TLR. TLR3 recognizes viral double stranded RNA whereas TLR7 and TLR8 detect viral single stranded RNA (16-18). TLR9 recognizes DNA containing unmethylated cytosine-guanosine (CpG) motifs common in bacteria (19). Recently, a specific sequence of 13 nucleotides present in bacterial 23S ribosomal RNA, was shown to release immune response via the murine TLR13 (20). Upon ligand binding the TLRs dimerize as homodimer or heterodimer and two TIR domains are brought in close proximity. A subsequent conformational change of the TIR domains provides a platform to recruit cytosolic adaptor proteins (21). TLR downstream signaling acts using two distinct adaptor molecules, namely TRIF and MyD88 (FIG. 1). Most of the nucleic acids sensing TLRs rely on MyD88 as adaptor molecule and thus

activation of IRF7 and Nf-κB. TLR3 is an exception and makes use of TRIF for downstream signaling resulting in IRF3 and NF-κB activation (22).

1.3.2 NOD-LIKE RECEPTORS

In contrast to TLRs, Nucleotide-binding and oligomerization domain (NOD) -like receptors (NLRs) recognize PAMPs localized in the cytoplasm. NLRs comprise either an N-terminal Caspase Activation and Recruitment Domain (CARD) or a pyrin domain (PYD) for signal transduction, a central nucleotide-binding and oligomerization domain (NOD/ NACHT) for ATP-dependent oligomerization and C-terminal leucine-rich repeats (LRRs) for ligand recognition. NLR protein's architecture resembles the one from plant R-proteins, which are implied in plant cell defense against microbes (23). There are 23 human NLRs that recognize a variety of PAMPs, including nucleic acids, deriving from intracellular pathogens (24). Upon ligand recognition via LRRs as with TLRs, NLRs undergo conformational changes exposing the N-terminal protein-protein interaction domain for downstream signaling. Some NLRs (NLRP1, NLRP3, NLRP6, NLRP7, NLRP12, NLRC4 and NAIP) are then able to build a platform in order to recruit and activate the inflammatory protease caspase-1. Active caspase-1 then cleaves the pro forms of the cytokines IL-1β and IL-18 leading to an inflammatory form of cell death, called pyroptosis (25).

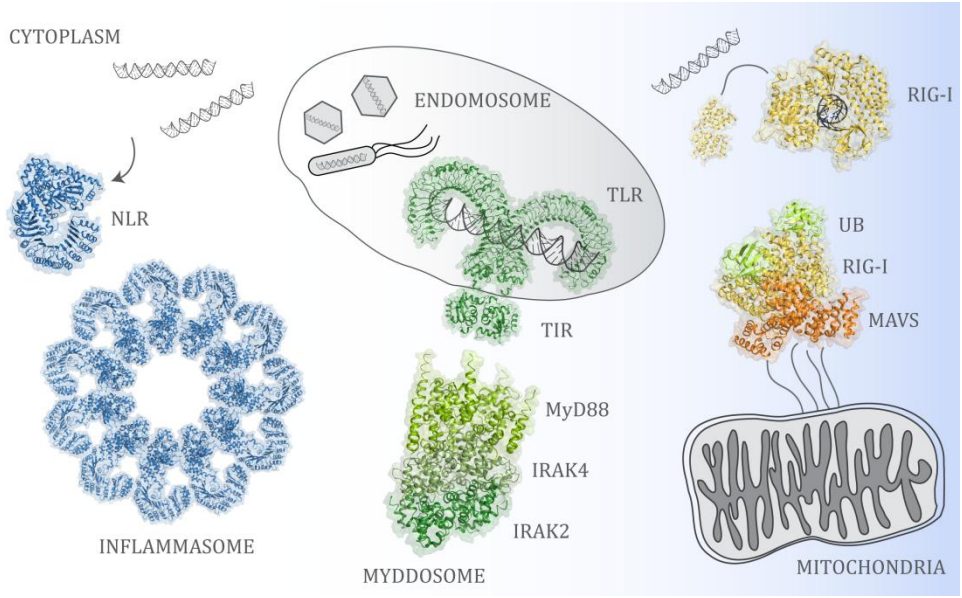


FIGURE 1: PATHOGEN RECOGNITION RECEPTORS SIGNALING I

Nucleic acids sensing pathogen recognition receptors are depicted along with their adaptor molecules required for downstream signaling. Only domains of the proteins where structural information is available are depicted. NLR (PDB: 4KXF) is shown in its auto-repressed state. NLRs are known to oligomerize to form an inflammasome structure (PDB: 3JBL) after activation with the respective PAMP. TLR (PDB: 3CIY) sense nucleic acids in the lumen of endosomes and activation dimerizes the cytoplasmic TIR (PDB: 4OM7) domains in the cytoplasm. This platform then recruits adaptor proteins MyD88, IRAK4 and IRAK2 that form the Myddosome (PDB: 3MOP). RIG-I (PDB: 5E3H) representing the RLRs, releases its CARDs (PDB: 4A2Q) after nucleic acid recognition and these oligomerize with the CARD of adaptor protein MAVS (PDB: 4P4H). Polyubiquitin chains stabilize these interactions forming a helical assembly resembling the Myddosome.

Out of these inflammasome constituting NLRs, NLRP3 has been reported to be activated, among others, upon release of oxidized mitochondrial DNA in the cytosol (26). As recently shown with the NLRC4 and NAIP2 inflammasome cryogenic electron microscopy structure, NLRP3 might assemble in a similar fashion after activation (FIG. 1) (27,28). Further, NLRX1, another NLR-family member with a unique mitochondrial targeting sequence, was revealed to directly interact with RNA ligands with its C-terminal LRR domain (29). This results in subsequent NLRX1-mediated reactive oxygen species (ROS) activation. The structural basis for ligand recognition by NLRs is poorly understood and it is still unclear if NLRs, except for NLRX1, directly bind nucleic acids or if adaptor molecules are required.

1.3.3 RIG-I-LIKE RECEPTORS

It was known for a long time that infection with RNA viruses, such as influenza and hepatitis C viruses, triggers a strong production of type I IFNs. This response was demonstrated in crucial studies to be independent from TLR signaling (30). Viral RNAs in the cytoplasm are recognized by the RIG-I-like Receptor (RLR) family, named after its founding member retinoic acid inducible gene-I (RIG-I). RIG-I was discovered in a function-based cDNA library screen in 2004 searching for proteins able to induce expression of an IFNβ promoter reporter gene. Together with its homologs melanoma differentiation associated gene 5 (MDA5) and laboratory of genetics and physiology 2 (LGP2), RIG-I was identified as the sensor of cytosolic viral RNA (31,32). RLRs are members of the superfamily II (SF2) of ATPases, helicases or nucleic acid translocases. RLR family members have a conserved domain architecture consisting of a central DExD/H-box helicase core with two helicase domains (Hel-1 and Hel-2) and an additional specific insertion within Hel-2 (Hel-2i). The C-terminal domain (CTD) confers part of the ligand specificity and is adjacent to Hel-2 by a helical extension named pincer domain. At the N-termini of RIG-I and MDA5, excluding LGP2, two tandem caspase activation and recruitment domains (CARDs) mediate signaling.

RIG-I and MDA5 distinguish between different kinds of RNA ligand and this is reflected in the distinct immune responses against specific virus families. RIG-I detects members of Paramyxoviridae, Flaviviridae and Rhabdoviridae family, while MDA5 responds to viruses of the Picornaviridae family. Among others Reovirus, Dengue and West Nile Virus (WNV) are detected by RIG-I as well as MDA5 (33-36). *In vitro*, the synthetic analog RNA (polyriboinosinic:polyribocytidylic acid) poly (I:C) elicits immune responses by RIG-I and MDA5. While MDA5 recognizes longer fragments >4 kb, RIG-I has a preference for shorter fragments after enzymatic digestion with ca. 300 bp (37). dsRNA containing a 5´ tri- (or di-) phosphate and lacking a 7-methylguanosine cap structure was identified as being the natural ligand of RIG-I (38-40). Further studies, demonstrated that RIG-I indeed is capable of binding the cap structure, consisting of a 5´ triphosphate 5´-5´ linked to a guanosine methylated at N_7. However, endogenous mRNA bearing a N_1-2´O-methyl group overcomes the risk of triphosphate recognition by RIG-I in the cytosol (41,42). Further, RIG-I activation was shown to additionally require base-paired RNA stretches such as panhandle-like structures as they occur in genomes of negative-strand ssRNA viruses (43-45). Among viral tri-phosphorylated RNAs, RIG-I is also able to recognize short RNA generated by RNAse L consisting of a double

strand stem together with a 5´hydroxyl (5´-OH) and a 3´-monophosphoryl group (46). In the case of MDA5, less is known about the ligand. Higher-order structured RNA containing ss and dsRNA arising during viral replication/ transcription resulted in activation of MDA5. These findings suggested that MDA5 activation requires not only long dsRNA, but rather a RNA forming web-like structures (47). LGP2 was shown to have the highest RNA binding affinity among RLRs, independently of length and 5´-triphosphate (48-50). Due to the lack of CARDs, LGP2 is not able of direct downstream signaling. A regulatory role in MDA5 and RIG-I dependent signaling, including synergistic effects with MDA5 leading to enhanced antiviral signaling, were assigned to LGP2 (51-53).

RIG-I and MDA5 share the same downstream signaling mechanism. In the absence of a RNA ligand both occur in an auto-inhibited conformation. Upon recognition of dsRNA, they undergo conformational changes exposing the N-terminal tandem CARDs (54-58). These are then accessible for E3 ligases TRIM25 and Riplet to modify them with K63-linked polyubiquitin chains (59,60). After the CARDs are released, RIG-I as well as MDA5 subsequently oligomerize along the RNA (61). Oligomerization of the CARDs provides a platform enhancing further interactions with the CARD of the downstream effector protein mitochondrial antiviral-signaling adaptor protein MAVS (also referred to as IPS-1, Cardif, and VISA) (FIG. 1) (62-65). Along with the N-terminal single CARD domain, MAVS has a proline-rich domain and a transmembrane domain at the C-terminus, which anchors MAVS in the mitochondrial outer membrane. Upon homotypic CARD-CARD interaction, MAVS undergoes massive aggregation on the mitochondrial membrane and oligomerizes into long helical filaments (66-68). MAVS prion-like aggregates function as a seed that leads to a robust amplification of the signal triggering activation of IRF3. The additionally K63 polyubiquitin chains on the RIG-I CARDs, have an increased affinity to the MAVS CARD and are therefore even more potent activators (69). MAVS oligomerization in turn recruits and activates through distinct TRAF-binding motifs multiple E3 ligases such as TRAF2, TRAF3, TRAF5 and TRAF6 that synthesize polyubiquitin chains (70-73). NEMO, the regulatory subunit of the IKK and TBK1, serves as a sensor for polyubiquitin chains (74,75). IKK and TBK1 complexes are recruited by NEMO to the MAVS polymers where these kinases phosphorylate IκBα and IRF3 resulting in the induction of type I interferons and other proinflammatory cytokines.

RNA binding modes for MDA5 and RIG-I, structural basis of MAVS helical filament activation and RIG-I's ATPase role in discrimination of self and non-self RNA will be described later on in detail.

1.3.4 OAS PROTEINS

Before the RLR's classification as PRRs for cytosolic dsRNA, in 1987 first antiviral activity of 2´-5´- oligoadenylate synthetases (OAS) was demonstrated (76). OAS1 dramatically reduced viral titers when transfected and overexpressed in cells infected with Encephalomyocarditis virus (EMCV). OAS proteins are constitutively expressed at low levels in the cell and serve as combined IFN-induced PRRs and antiviral protein (77). Upon sensing dsRNA, OAS proteins have the unique property to catalyze the conversion of ATP into 2´-5´-linked oligoadenylates

(78). These 2´-5´-linked oligoadenylates subsequently bind to the latent ribonuclease RNase L leading to dimerization and the activation of a viral as well as cellular RNA degradation pathway already described in the late 1970s (FIG. 2) (79-83). Protein translation and viral replication are thus inhibited in virus infected cells. Cleavage products of RNase L further augment antiviral response through recognition by PRRs, such as RIG-I, followed by induction of type I IFN production (46,84). However, RNase L cleavage products characterized by a 3´-monophosphate seem to have a minor contribution in RIG-I activation compared to genomic RNAs generated during viral replication (85).

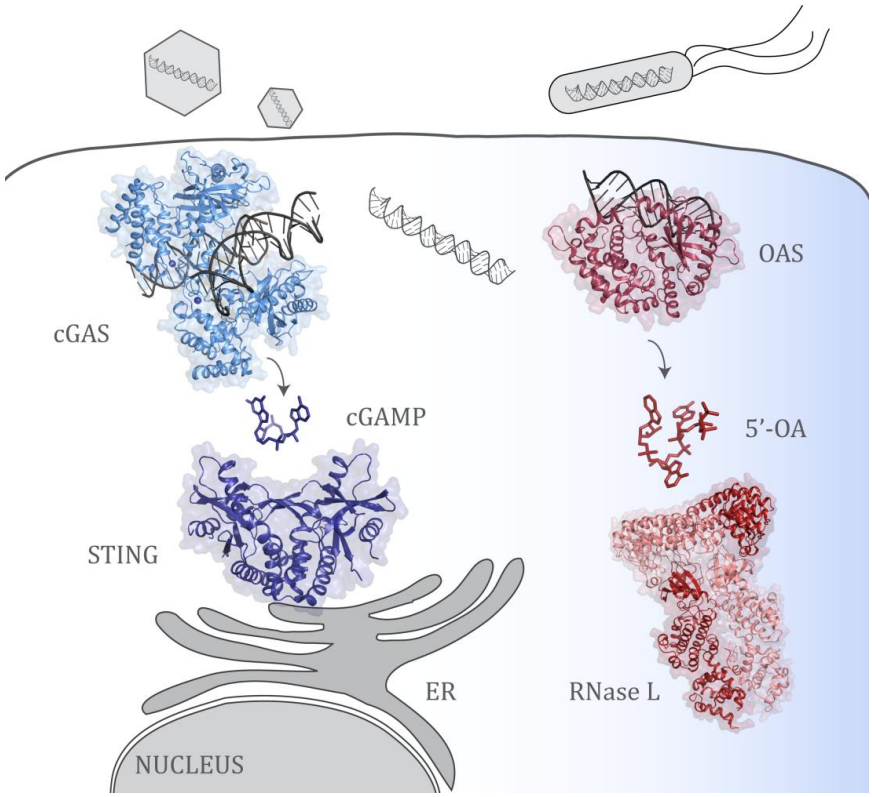


FIGURE 2: PATHOGEN RECOGNITION RECEPTOR SIGNALING II

Nucleic acids sensing pathogen recognition receptors are depicted along with their synthesized 2´-5´-linked second messengers and adaptor molecules required for downstream signaling. Only domains of the proteins where structural information is available are depicted. cGAS (PDB: 4LEY) dimerizes upon dsDNA recognition and synthesizes the second messenger cGAMP, which is then sensed by the adaptor protein STING (PDB: 4KSY) followed by dimerization and subsequent activation of downstream signaling. OAS1 (PDB: 4IG8) produces 2´-5´-oligoadenylates in presence of dsRNA in the cytosol. This second messenger is then recognized by the downstream adaptor RNase L (PDB: 4OAU). cGAS and OAS proteins form a conserved class of pathogen recognition receptor synthesizing small molecules containing a 2´-5´-phosphodiester linkage for downstream signaling.

There are four OAS genes in humans OAS1, OAS2, OAS3 and OASL that share considerable homology (86). The basal OAS unit is repeated once in OAS1 and OASL, twice in OAS2 and thrice in OAS3. Each of these OAS proteins has unique physiological functions. Oligomerization of OAS1 and OAS2 in its catalytically active form is mediated by a tripeptide motif (CFK) within the OAS domain (87). OAS3 and OASL lack this motif and thus function as monomers. Copies of OAS units are catalytically inactive and in the case of OASL mutations at key residues in the active site completely abolish its catalytic function. C-termini of OAS proteins are variable and OASL encodes a tandem ubiquitin-like domain (88). Besides lacking catalytic activity, OASL retains its antiviral properties suggesting an RNase L independent antiviral effect of the OAS family (89-92). OASL antiviral activity relies on the feature of its tandem ubiquitin-like domain to substitute for K63-linked poly-ubiquitin interacting with the CARDs of RIG-I and thereby enhancing RIG-I signaling (93,94). Furthermore, structural studies revealed the RNA binding groove of OASL and demonstrated that the ability of OASL to bind RNA is required for amplifying antiviral response via RIG-I.

1.3.5 cGAS-STING PATHWAY

Long before DNA was identified as the genetic material it was known to elicit a potent immune response (95). DNA was shown to induce type I IFN through the protein stimulator of interferon genes (STING, also termed as TMEM173, MPYS, MITA and ERIS) (96-98). STING has an N-terminal transmembrane domain that confines it to the endoplasmatic reticulum (ER), a cytoplasmic soluble cyclic-di-nucleotide (CDN) binding domain and a C-terminal tail (CTT). Bacterial CDNs, such as c-di-GMP were shown to directly activate STING (99,100). Upon activation STING oligomerizes into a supramolecular complex and translocates from the ER to a perinuclear compartment, where it activates NF-κB and IRF3 through interactions of its CTT with the kinases IKK and TBK1 (101).

However, STING is not able to directly bind DNA and therefore an additional host protein is required to function as a DNA sensor (102). Recently the search for an unambiguous and non-redundant cytosolic DNA sensor acting upstream of STING was successful. The cyclic GMP-AMP synthase (cGAS) was identified as the major PRR for cytosolic DNA through biochemical purification coupled with quantitative mass spectrometry (103,104). cGAS contains an N-terminal unstructured stretch followed by an NTase fold domain, harboring the conserved motif found in the nucleotidyltransferase (NTase) fold protein family and partially overlapping with the C-terminal male abnormal 21 (MAB21) domain. When DNA is sensed, cGAS binds ATP as well as GTP and produces a novel second messenger: a cyclic GMP-AMP dinucleotide with a noncanonical 2′-5′ and a 3′-5′ mixed phosphodiester linkage (FIG. 2) (105,106). This CDN is in turn recognized by STING inducing a closed conformation of the STING dimer with four antiparallel β-sheet strands forming a lid above the bound CDN (107,108). STING is more responsive to 2′-3′-cGAMP than canonical bacterial CDNs harboring 3′-5′ linkages and therefore cGAMP is a more potent stimulator of the IFN pathway (105,109). Structural studies revealed that cGAMP binds 2.5 Å deeper in the cleft of the STING dimer than c-di-GMP inducing closure of the dimer by 20 Å and lid formation (110). cGAMP was further shown to promote antiviral response in a transcription-independent horizontal

manner bypassing type I IFN signaling. The second messenger is transferred from infected cells via gap junctions inducing STING activation in neighboring cells (111). In addition, viral particles, such as lentivirus or herpes virus, where shown to incorporate cGAMP, when produced in cGAS-expressing cells (112,113). In newly infected cells, cGAMP then activates STING spreading and accelerating immune response against the progeny virus.

Further experiments showed that, while TLR9$^{-/-}$ and MyD88$^{-/-}$ mice cleared the herpes simplex virus 1 (HSV1) infection as effectively as the wild type, cGAS$^{-/-}$ mice were more susceptible to lethal infection due to failure in type I IFN production (114). cGAS was thus unveiled to unequivocally be the general PRR for dsDNA. Moreover, cGAS acts as a retroviral sensor, able of detecting reverse-transcribed viral cDNA from HIV and other retroviruses (115). Concerning the immune response against RNA viruses, cGAS plays an indirect role by contributing to tonic type I IFN levels, facilitating rapid immune response. This low constitutive type I IFN response is triggered by cGAS ability to recognize host DNA released from cells damaged or killed by the virus (116). Another event where cGAS was shown to recognize self-DNA is after infection with HSV1, a dsDNA virus. This infection induces aberrant mitochondrial DNA packaging leading to release of mtDNA into the cytosol where it is recognized by cGAS and activates an antiviral response (117). This cell-intrinsic signal of infection can be mediated by mitochondria just as it can be inactivated in a caspase-dependent manner (118). The cGAS-STING pathway was further shown to be antagonized by oncogenes of DNA tumor viruses, such as human papillomavirus (HPV) and adenovirus (119). These findings explain why several immortalized cell lines have suppressed IFN signaling and suggest that a principal function of viral oncogenes, which shaped the virus' evolution, is to inhibit the innate immune response.

Structural insights into DNA recognition by cGAS, catalytic reaction mechanism of cGAMP synthesis and homology to the OAS family will be presented in detail further in this work.

1.3.6 OTHER DNA SENSING RECEPTORS

Before cGAS was acknowledged as the central PRR for cytosolic dsDNA, several studies were conducted revealing other potential nucleic acids sensing receptors. The first protein to be proposed as a cytosolic DNA sensor was the DNA-dependent activator of IRFs (DAI, also termed ZBP1 or DLM1) (120). DAI contains DNA-binding domains (Z-α and Z-β) and knockdown studies resulted in a modest defect in type I INF production (121). However, no physical or genetic link to STING was presented and DAI knockout mice experiments showed that DAI is not indispensable in order to mount a type I IFN response (122).

Another superfamily of potential DNA sensors is the PYHIN protein family also termed as AIM2-like receptors (ALR). ALRs were shown to sense HSV-1 and Kaposi's sarcoma virus DNA in the nucleus, to immunoprecipitate with STING in dsDNA-stimulated cells and to function in the inflammasome pathway (123-127). Members like AIM2 and IFI16/ p204 possess C-terminal hemopoietic expression, interferon-inducible, nuclear localization (HIN) domains that interact with B-form DNA, and an N-terminal effector PYRIN domain (PYD). The

PYDs from AIM2 and IFI16 were shown to form filaments which then engage with the apoptosis-associated speck-like protein containing a CARD (ASC), leading to inflammasome assembly and subsequent proteolytic cleavage of inactive precursors of IL-1β and IL-18 (128,129).

DDX41, member of the superfamily of DExD/ H-box RNA helicases, is another interesting candidate as additional DNA sensor. Interferon response to dsDNA as well as to the DNA virus HSV-1 was affected when DDX41 was knocked down. DDX41 was shown to bind DNA instead of poly (I:C) despite its homology to RNA helicases (130). And further studies pointed out co-localization of DDX41 with STING and ability to directly sense bacterial CDNs (131).

Another link was proposed between DNA damage repair and DNA-induced immune response. Other putative DNA sensors, such as DNA-dependent protein kinase (DNA-PK) and meiotic recombination 11 homolog A (MRE11) were shown to bind DNA and be required for type I IFN response via STING (132-134). Nevertheless, a major issue concerning this search for further cytosolic DNA sensors is that it relies on assays such as immunoprecipitation, immunofluorescence and RNAi knockdown experiments which can lead to indirect evidences and non conclusive results. Genetic evidence demonstrating requirement for the sensor and direct interactions with STING together with the more conclusive gene knockout experiments rather than RNAi knockdown would provide definitive results (102).

1.4 cGAS – THE MAJOR SENSOR FOR CYTOSOLIC DNA

1.4.1 STRUCTURAL STUDIES ON cGAS DNA SENSING AND ACTIVATION

After the discovery of cGAS as the major DNA sensor, several crystal structures were published of cGAS catalytic MAB21 domain in complex with DNA, substrate nucleotides, linear intermediates and the final product cyclic G(2´-5´)pA(3´-5´)p (106,135-137). cGAS has a bilobal structure with mixed α/β topology and separated by a deep cleft harboring the active site (FIG. 3). Lobe 1 comprises the N-terminal NTase subdomain with central highly twisted β-sheets (β1-β8) flanked by two long α-helices (αA-αB). The catalytic triad (*hsc*GAS E225, D227 and D319/ *ssc*GAS E200, D202 and D296) responsible for Mg^{2+} coordination in other NTase superfamily members is conserved and located on β-sheets β1 and the adjacent β6. The C-terminal lobe 2 is composed of a conical four-helix bundle (αE-αH). This structure is stabilized by a long N-terminal 'spine' helix (α1) that spans across both lobes and an additional central linker region containing two helices (αC, αD) and a long active site loop between αA and β1. The loop connecting αD and αE contains a highly conserved set of a histidine and cysteines (*hsc*GAS H390, C396, C397 and C404 /*ssc*GAS H367, C373, C374 and C381) that coordinate a Zn^{2+}ion (Zn-thumb). Inactive cGAS is monomeric in solution and the active site is partially unstructured with B-factors of lobe 1 being twice as high as the ones for the less flexible lobe 2 (109). In the absence of dsDNA, the active site is not properly aligned for coordination of nucleotide substrates. Structures are available of inactive NTases with not

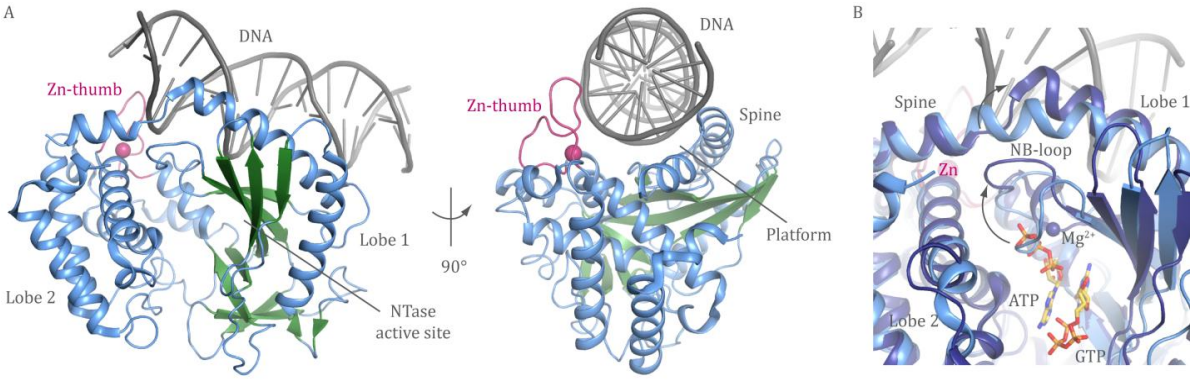


FIGURE 3: DNA RECOGNITION BY CYTOSOLIC DNA SENSOR cGAS

A. Front and side views of cGAS' MAB21-domain (PDB: 4LEY) structure in complex with dsDNA (grey). The model is shown as cartoon representation with annotated domains. DNA binds along the 'platform' between the spine helix and the Zn-thumb (pink). B. Close up view of aligned apo cGAS (light blue, PDB: 4K8V) and substrate- and DNA-bound activated cGAS (dark blue, PDB: 4KB6). The structural switch induced by the helix break on the spine, allows for closure of the active site cleft and proper residue alignment on the nucleotide-binding loop.

correctly coordinated nucleotides bound to the active site, such as the cGAS structure in complex with UTP (135,138).

On the opposite site of the NTase active site, cGAS displays a highly positively charged molecular 'platform'. Structures of cGAS bound to dsDNA revealed that this 'platform' binds to dsDNA in a sequence-independent manner where binding is mainly mediated by extensive electrostatic interactions and hydrogen bonds with the phosphate backbone (106,135). In addition two arginine fingers (*ss*cGAS R150, R192/ *hs*cGAS R176, H217) reach into the minor groove of the DNA. The protruding loop stabilized by the Zn^{2+} ion is rigid upon dsDNA binding and serves as a molecular ruler to specifically distinguish between B-form DNA and A-form RNA. Binding of B-form DNA to cGAS induces a helix break in the 'spine' that closes lobe 1 and 2 rearranging the active site loops in order to allow coordination of two Mg^{2+} ions and the substrate nucleotides. Further studies using analytical ultracentrifugation and small-angle X-ray scattering experiments revealed that cGAS forms a 2:2 complex with dsDNA in solution instead of the 1:1 complex (FIG. 4) (109,139). Dimerization further stabilizes the active conformation in which two dsDNA molecules are sandwiched between two cGAS protomers. Besides the 'platform' denoted as site A, a secondary DNA binding site B and the Zn-thumb, were shown to be important for DNA induced oligomerization of cGAS. Mutations at site B reduced DNA binding as well as higher-order complex formation more drastically than mutations at site A, suggesting that site B plays a role in cooperative DNA binding (109). In its inactive conformation cGAS NTase domain is locally destabilized. During DNA binding, site A is responsible for triggering the structural switch, while site B contributes to cooperative binding. The resulting crosslinking of the cGAS dimer restricts previous structural flexibility which is supported by the uniform distribution of B-factors in the complex structures (109).

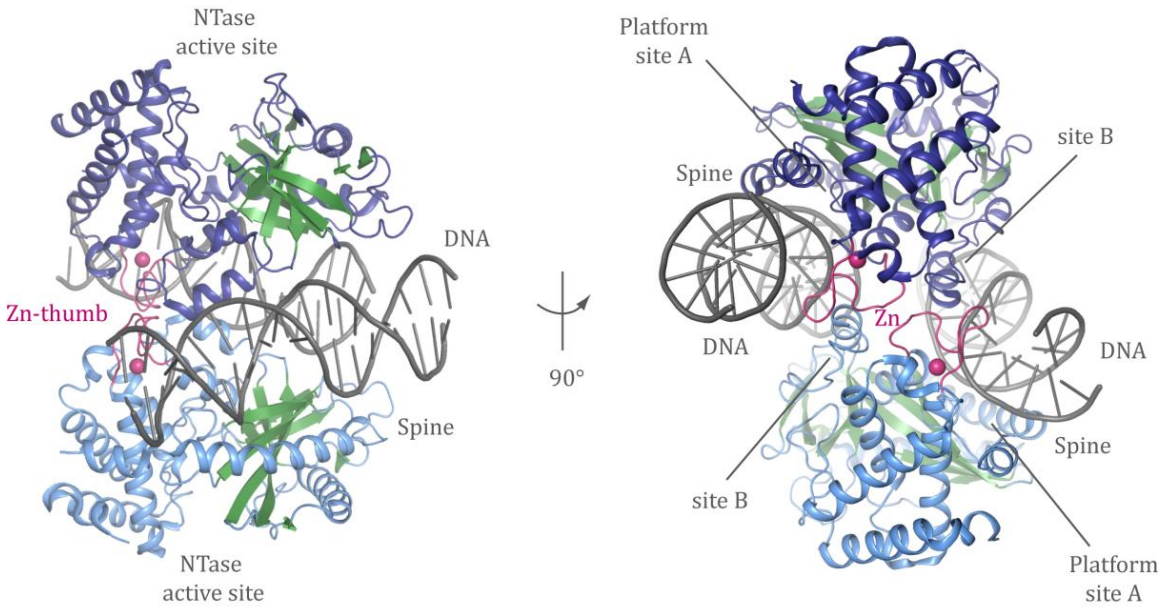


FIGURE 4: cGAS OLIGOMERIZES UPON DNA BINDING REVEALING A SECOND DNA BINDNG SITE

Top and front views of the MAB21 domain of cGAS in a 2:2 cGAS:dsDNA complex in the crystallographic asymmetric unit (PDB: 4LEY). Both 18bp dsDNA strands are depicted in grey, cGAS protomers in light and dark blue respectively and the Zn-thumbs in pink. Both DNA binding sites A (platform) and B are marked.

1.4.2 cGAS PRODUCES 2´-5´, 3´-5´ CYCLIC GMP-AMP

Regarding the catalytic mechanism of cyclic G(2´-5´)pA(3´-5´)p synthesis, the *ssc*GAS complex structure bound to DNA, ATP and GTP using transferase-trapping active site mutations provided an insight into the first catalytic step leading to the formation of linear intermediate pppGp(2´-5´)A (FIG. 5) (135). GTP is bound to the N-terminal acceptor site, while ATP sits in the C-terminal donor site with the two bases partially stacking to each other in an approximately 90° rotated orientation. In this first catalytic step the triphosphate from ATP is coordinated by the two Mg^{2+} ions in a way that the α-phosphate can be attacked by the 2´-OH of GTP. In addition D296 polarizes the attacking nucleophilic 2´-OH from GTP, Y413 stacks the adenine base and E360 holds the donor ribose in place.

Subsequent structural studies of wild type cGAS revealed the cyclization reaction during the second catalytic step. When comparing structures of cGAS with the linear product pG(2´-5´)pA, it is obvious that the cyclic product G(2´-5´)pA(3´-5´)p is in the opposite orientation. In the structure with bound linear intermediate the acceptor guanine stacks with *ssc*GAS Y413/ *hsc*GAS Y436 whereas the structure with bound cGAMP shows how the donor adenine stacks with *ssc*GAS Y413/ *hsc*GAS Y436 (106). This suggests that the linear intermediate pppGp(2´-5´)A has to flip-over and rebind in the reversed order into the active site for the cyclization reaction to take place. GTP then occupies the donor pocket and 2´-5´-linked AMP is in the acceptor site. Formation of the 3´-5´ phosphodiester bond takes place by an additional

nucleophilic attack of the AMP 3´-OH onto the α-phosphate of GTP, which is consent with the general nucleotidyltransferase reaction mechanism (140,141).

Different structures of cGAS with cyclic G(2´-5´)pA(3´-5´)p bound show that after 3´-5´ phosphodiester bond formation a second flip-over takes place just before the product is released (106,109). cGAS active site has distinct recognition elements that enable differentiation between guanine and adenine. Whereas the donor binding site is promiscuous and adenine can even be replaced by a pyrimidine, the guanine base forms a network of intermolecular hydrogen bonds with its Watson-crick as well as Hoogsteen edges with the acceptor binding site (106). Thus, cGAMP is bound in a unique orientation in the cGAS active site pocket. In the case of the linear intermediate pppGp(2´-5´)A, the triphosphate seems to be capable of overruling the donor acceptor specificities. The two step synthesis of cyclic G(2´-5´)pA(3´-5´)p by cGAS is distinct from the one of bacterial diadenylate (DAC-domain) and diguanylate cyclases (GGDEF-domain) (142,143). These bacterial enzymes are dimers or oligomers and each active site binds one ATP or GTP. Two opposing active sites carry out two simultaneous transferase reactions to form a CDN with two 3´-5´ phosphodiester linkages.

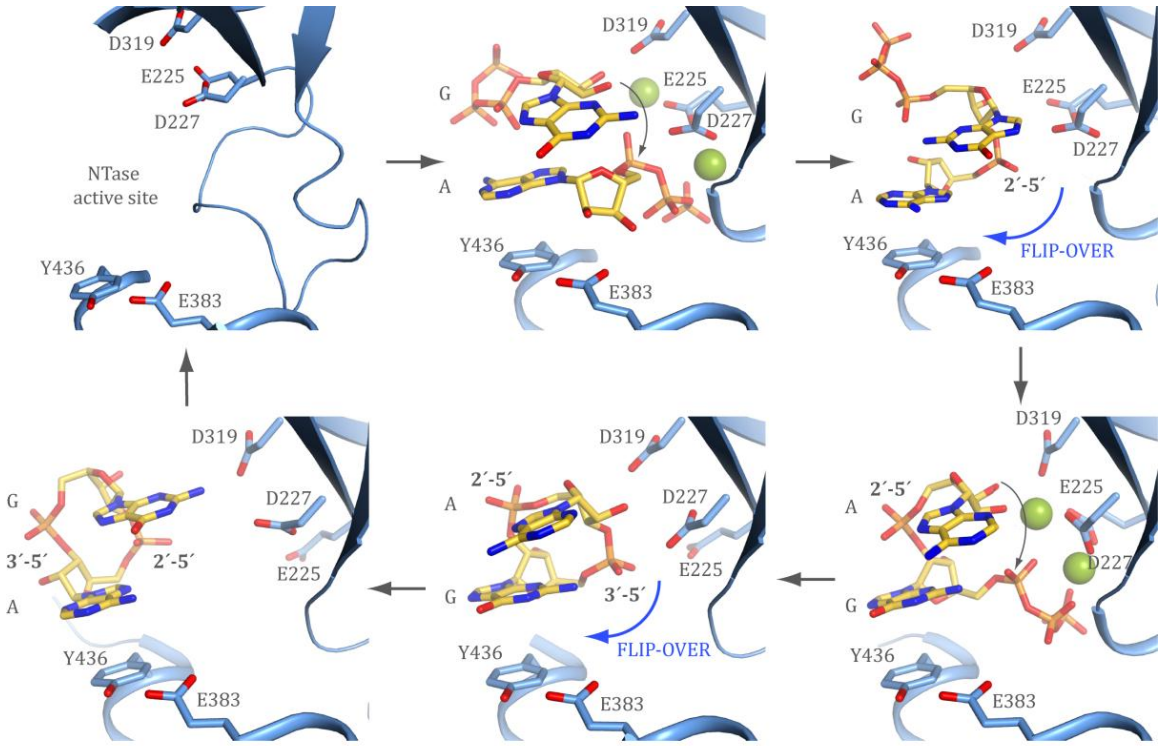


FIGURE 5: CURRENT MODEL OF THE CATALYTIC MECHANISM OF CYCLIC G(2´-5´)pA(3´-5´)p SYNTHESIS BY cGAS

Schematic representation of the current model involving a two-step generation of cyclic G(2´-5´)pA(3´-5´)p within the single catalytic pocket of cGAS. cGAS binds ATP in the donor binding site and GTP in the acceptor binding site and the first step involves the formation of a 5'-pppGpA linear intermediate which is flipped-over in order to allow the second reaction step where the cyclic product G(2´-5´)pA(3´-5´)p is generated. In order to facilitate product release, a second flip-over occurs.

1.4.3 cGAS - OAS FAMILY

Recent studies revealed that OAS proteins and cGAS form a new family of template-independent nucleotidyltransferases that act as cytosolic nucleic acids sensing PRRs (106,135-137). OAS proteins and cGAS are not phylogenetically related (144). However, they share some functional properties, such as nucleic acid binding, synthesis of a second messenger containing a noncanonical 2´-5´ phosphodiester bond and the ability of viral inhibition. Regarding the structural similarities, cGAS and OAS1 have a common overall fold, conserved active site residues and share activation induced rearrangements of lobe I and II, as well as structural elements such as the 'platform' and the 'spine' helix (145).

While OAS1 binds ssRNA, cGAS was shown to bind ssDNA or even dsRNA *in vitro* (135). Nevertheless, it is not sufficient to trigger the structural switch required for activation. Thus it seems that OAS1 and cGAS require the double-strand topology to distinguish between A-form RNA and B-form DNA. cGAS is activated by B-form DNA interactions with the 'platform', Zn-thumb and site B of the other cGAS protomer, whereas OAS1 remains monomeric and requires A-form RNA binding contacting two minor groove segments separated by 30 Å (135,138). Modeling studies of dsRNA and dsDNA into the cGAS apo structure revealed that DNA clashes with the protruded activation loop forcing it to move inward and subsequently rearranging the active site (139). In the case of RNA binding, modeling showed that the same activation loop inserts into the major groove devoid of conformational changes. When comparing the available crystal structures, it becomes clear that the cGAS apo form is much closer to its active state than in the case of OAS1. Differences in the N-terminal lobe, explain the more extensive conformational changes of OAS1 upon activation (146).

Taking into account the active site specificities, cGAS as well as OAS1 have promiscuous donor binding sites, which were shown to accept any triphosphate nucleotide. In contrast orientation of the substrate by the acceptor site, determines the 2´-specificity of the product formation. One major difference between OAS and cGAS is that cGAS catalyses the synthesis of CDNs, while OAS1 forms linear 2´-5´ oligoadenylate chains. OAS1 suppresses the flip-over of the intermediate pppA(2´-5´)pA by sterically blocking CDNs with helix αE (3_{10}3-α5) (FIG. 6) (146). Instead, this linear intermediate has to rebind

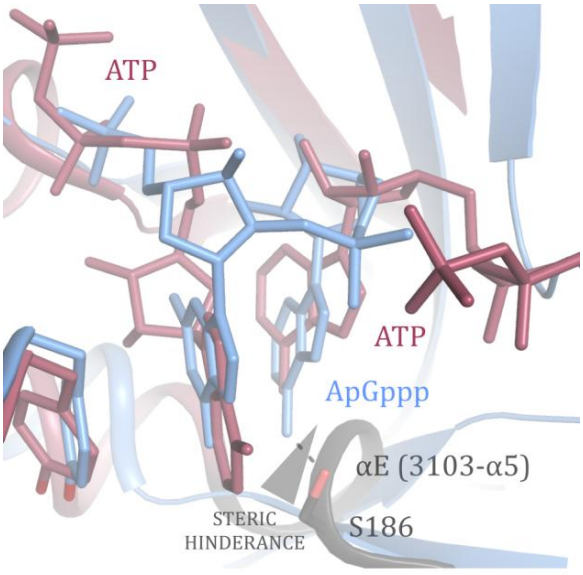


FIGURE 6: HELIX αE IN THE ACTIVE SITE OF OAS1 HINDERS CYCLIC PRODUCT FORMATION

Superposition of the active sites of cGAS (blue, PDB: 4K9A) bound to linear intermediate ApGppp and OAS1 (red, PDB: 4RWN) bound to two ATP molecules. OAS1 suppresses the flip-over of the linear intermediate by sterically blocking CDN coordination with S186 on helix αE.

to the acceptor site (145). Further steps in 2´-5´ oligoadenylate chain elongation use pppA(2´-5´)pA as an acceptor for following 2´-5´ phosphodiester linkages.

One further event underlying the evolutionary connection between OAS proteins and cGAS, is the extensive and common positive selection these enzymes experienced in order to evade inhibition by pathogens (144,147). Several amino acid substitutions were located at equivalent positions in the protein-nucleic acid interfaces comprising the 'spine' and the C-terminus of αE. This suggests that cGAS and OAS proteins adapted to evade common pathogen encoded inhibitors. In addition cGAS and OAS genes made use of alternate splicing or gene fusion events respectively, in order to drastically alter protein structure and thus evade recognition by certain pathogens.

1.5 MEMBERS OF THE NTASE FOLD PROTEIN FAMILY MAB21

Nucleotidyltransferases (NTases) constitute a large and diverse family of proteins having a variety of distinct functions in the cell. All NTases catalyze the transfer of a nucleoside monophosphate (NMP) from a nucleoside triphosphate (NTP) onto an acceptor hydroxyl group via elimination of diphosphate (148,149). As an acceptor a vast range of substrates are available from proteins, nucleic acids to small molecules. Besides displaying little sequence similarity, NTases share a common fold with three-stranded, mixed β-sheet flanked by four α-helices with αβαβαβα topology (150). The core NTase fold harbors common sequence patterns, that include the active site residues hG[GS], [DE]h[DE]h and h[DE]h (h indicates a hydrophobic amino acid) (148). Two aspartate or glutamate residues located on one core β-sheet are involved in the coordination of divalent ions essential for catalysis. The third carboxylate residue on the adjacent β-sheet is responsible for the activation of the acceptor hydroxyl group. On the nucleotide-binding loop (NB-loop), forming an α-helical turn, is the antecedent hG[GS] motif normally crucial in harboring the substrate within the active site.

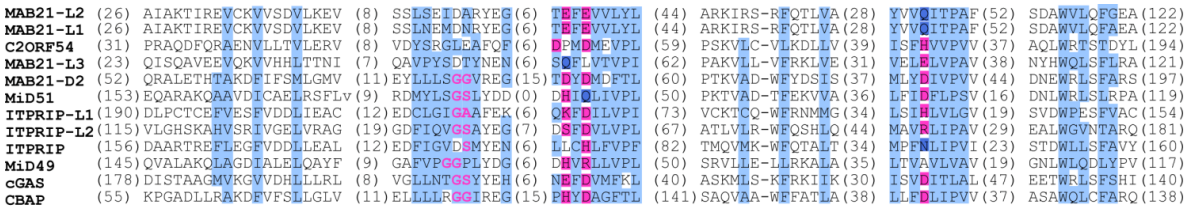


FIGURE 7: ALIGNMENT OF THE TWELVE NEWLY CLASSIFIED MAB21 FAMILY MEMBERS

Multiple sequence alignment for the NTase fold protein subfamily MAB21 (150). Only conserved regions of the fold core are shown. Conservation of residues is denoted with the following scheme: uncharged, highlighted in light blue; small, letters in pink; critical active site aspartate/ glutamate, highlighted in pink and critical inactive asparagine and glutamine, highlighted in dark blue.

A previous comprehensive review of the nucleotidyltransferase (NTase) fold protein family significantly expanded the number of putative NTases (150). Using top in line homology

detection methods combined with fold recognition programs, 25 novel superfamily members were identified in the human genome. Twelve of these newly classified members belong to the MAB21 family: MAB21L1, MAB21L2, MAB21L3 (C1ORF161), cGAS (MB21D1, C6ORF150), MB21D2 (C3ORF59), MiD51 (SMCR7L), MiD49 (SMCR7), CBAP (TMEM102), ITPRIP (KIAA1754, DANGER), ITPRIPL1, ITPRIPL2 and C2ORF54 (FIG. 7). All members possess an N-terminal NTase domain followed by C-terminal PAP/ OAS1 substrate binding domain. Biological functions of most of these NTases are unknown and some even partially lack the conserved catalytic residues and thus seem to lack enzymatic activity.

1.5.1 TWELVE NEW HUMAN NTASE FOLD PROTEINS- THE MAB21 FAMILY

Besides the best characterized MAB21 family member being cGAS (MB21D1), the MAB21 family was named after a conserved family of eukaryotic MAB21-like proteins (MAB21L1, MAB21L2 and MAB21L3) involved in cell fate specification (151). MAB21-like proteins will be described in detail later in this chapter. Regarding other members of this family, the common β chain associated protein (CBAP, TMEM102) was shown to play an important role in: pro-apoptotic signaling via interactions with the GM-CSF/ IL-3/ IL-5 receptor (152), chemokine-induced T-cell migration and adhesion (153) and thymocyte development through T-cell receptor signaling promoting negative selection (154). Another member termed DANGER (ITPRIP) is a membrane-associated protein predicted to contain a partial MAB21 domain. DANGER's MAB21 domain interacts with the inositol 1,4,5 triphosphate receptor (IP_3R) and co-localizes to membranes in the cell periphery, allosterically modulating the sensitivity of IP_3R to inhibition by Ca^{2+} (155). Further studies demonstrate that DANGER directly binds the calcium/ calmodulin-dependent death associated protein kinase (DAPK) inhibiting its catalytic activity and thus regulating cell death (156). DANGER deletion augments for instance neurotoxicity and therefore DANGER has the potential to serve as therapeutic target (156).

Mitochondrial dynamics proteins of 49 kDa and 51 kDa (MiD49/SMCR7 and MiD51/SMCR7L) share 42 % sequence similarity and were found in a random cellular localization screen where expression of MiD49 led to fused mitochondrial tubules (157). MiD proteins are anchored in the mitochondrial outer membrane with their N-terminal transmembrane domain and the C-terminal MAB21 domain is located in the cytosol. MiD49 and MiD51 were shown to participate in the mitochondrial fission machinery mediating recruitment of the GTPase dynamin-related protein 1 (Drp1) independently from other mitochondrial outer membrane proteins: mitochondrial fission protein 1 (Fis1) and mitochondrial fission factor (Mff) (158-160). MiD49 and MiD51 seem to be functionally redundant since changes in mitochondrial morphology were only observed when both MiDs were knocked down. Recently, first crystal structures of MiD proteins revealed ADP as essential co-factor of MiD51 for activation of Drp1 self-assembly (FIG. 8) (161,162). In the crystal structure ADP is bound in the catalytic cleft and stabilizes MiD51 devoid of conformational changes. Differently from

other NTases the catalytic residues H201 and Q203 rather directly interact with ADP than coordinate the Mg^{2+} ions. MiD51 displays no affinity for nucleotide triphosphates, and instead seems to prefer ADP or GDP suggesting that MiD51 is catalytically inactive. Intriguingly, MiD51 crystal structure seems to trap MiD51 in its active state displaying the helix break in the 'spine' and the α-turn in the NB-loop. In addition, MiD51 dimerizes using a distinct interaction interface from that known of other NTases. An important finding was that a surface loop (β4-α4) is required for Drp1 binding independently of MiD51 dimerization and ADP binding. Drp1 recruitment was abolished when a single residue R235 was mutated disrupting a salt bridge (R235-D249) located on this loop (161).

Purified MiD51 can recruit Drp1 to the mitochondrial outer membrane and is able of inhibiting Drp1 assembly. Only in the presence of ADP, Drp1 can be activated for fission, resulting in assembly into spirals and GTP hydrolysis. MiD49 lacks key residues (only H193 is present) and it is monomeric since the dimer interface is not conserved. The binding affinity for ADP is also lost. Instead, MiD49 has a modified binding pocket where the central β-sheet is repositioned and the flexible NB-loop sterically hinders co-factor binding in the active site (163). Other MAB21 family members such as MB21D2 or C2ORF54 have completely unknown function even though they harbor the conserved catalytic residues.

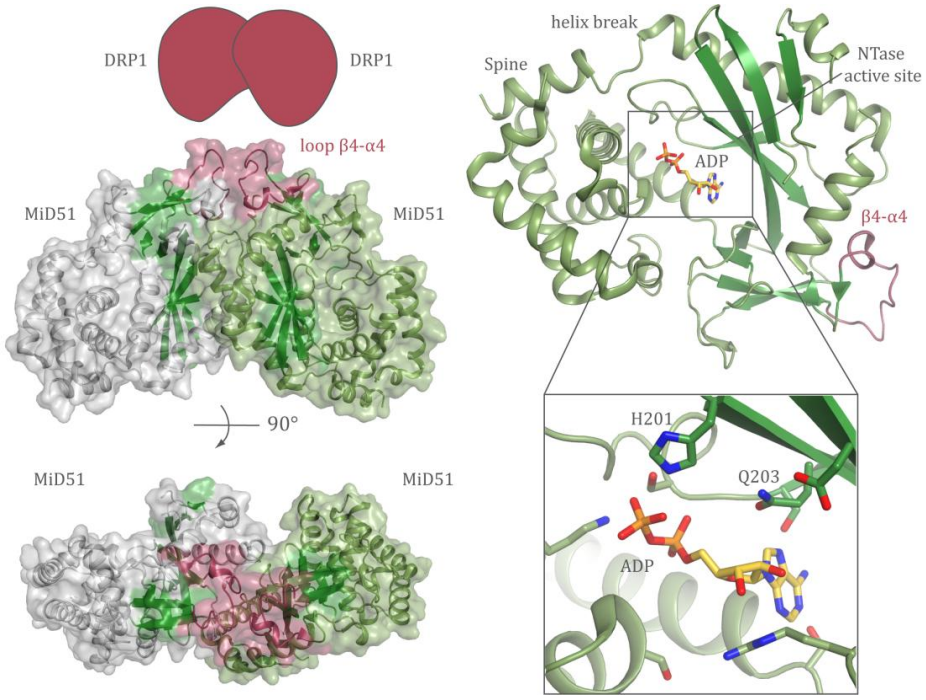


FIGURE 8: THE CYTOSOLIC DOMAIN OF MiD51 HAS A NUCLEOTIDYLTRANSFERASE FOLD AND BINDS ADP

Left panel: side and top views of the crystallized MiD51 (4OAG) dimer showing the dimer interface in the N-terminal domain. The surface loop β4-α4 required for Drp1 recruitment is depicted in red. Right panel: Crystal structure of MiD51 bound to ADP with a close up view of the active site. ADP (yellow) is directly coordinated by residues H201 and Q203.

1.5.2 MAB21 CELL FATE DETERMINING PROTEINS

The *mab21* (male abnormal 21) gene was first described in *Caenorhabditis elegans* as a cell fate determinant (164). Mutations in 6 genes were shown to alter the adult arrangement of sensory rays on the nematode's tail and four of them specifically affected ray development (165). Rays are a set of nine bilateral and symmetrical pair of protruding structures, each consisting of an epidermal structural cell, a neuron A and a neuron B cell. Each ray tip has a precise antero-posterior and dorso-ventral localization on the surface of the nematode. Mutations in the *mab21* gene result in loss of ray 6 morphology and lead to fusion of this ray with ray 4 (165). Other pleiotropic defects include abnormal locomotion, shorter body shape and reduced fecundity, suggesting that *mab21* function is not exclusively in the male tail. All effects of *mab21* mutations can be seen as posterior-to-anterior homeotic transformations. In addition, *mab21* mutations lead to the transformation of a neighboring hypodermal cell (T. apapa) into a neuroblast cell expressing ray sublineage and that generates a ray (165). Thus *mab21* mutants exhibit an additional tenth ray, termed ectopic ray, instead of 9 rays. Cell ablation experiments confirmed that *mab21* is able of acting cell autonomously when specifying the properties of sensory rays and acts cell non-autonomously in the case of the neuroblast cell fate choice (164). Epistatic analysis revealed that *mab21* acts downstream of the Sma/Mab pathway (TGF-β pathway in vertebrates) (166,167). A ligand for this pathway *dbl-1* (also termed *cet-1*, a homolog of vertebrate bone morphogenetic protein, BMP) acts as general dorsalizing factor in male ray patterning and negatively regulates *mab21*. Further characterization of MAB21 function in *C. elegans* so far, involves the interaction and co-expression in sensory rays with SIN-3, a key component of a histone deacetylase associated co-repressor complex. Regarding developmental and post-embryonic expression patterns, MAB21 fused to GFP was first observed in embryos at the beginning of gastrulation and could be traced in larval and adult nematodes in hypodermal cells, neuronal cells, neurons along the ventral nerve cord, body wall, muscles and ray cells (151).

Two highly conserved vertebrate MAB21 orthologs were identified and designated MAB21-like 1 (MAB21L1) and MAB21-like 2 (MAB21L2) (168-170). The *mab21* gene family exhibits striking phylogenetic conservation and vertebrate orthologs are 94 % identical in amino acid sequence. MAB21L1 and MAB21L2 possibly aroused from an ancestral gene duplication event prior to vertebrate divergence, as they are equidistant from their *C. elegans* homolog. The ancestral *mab21* gene was embedded in an intron of *Lrba/ Nbea* before metazoan divergence leading to a conserved nested gene pair (*MAB21L2-Lrba* and *MAB21L1-Nbea*) (171). Several studies in the different model organisms *Xenopus*, zebrafish and mouse revealed that both genes have similar expression patterns in the developing eye, midbrain, neural tube, branchial arches and limb bud (172-176). MAB21-like proteins have widespread and dynamic expression patterns during embryonic development. And yet these expression patterns are still similar in different vertebrate model organisms, suggesting that these proteins act in concert during embryonic tissue patterning and organogenesis. In addition, *MAB21-like* genes maintain expression in mature tissues, such as postnatal cerebellum and eye, and have therefore not an exclusively developmental role (172,177). Alongside with

overlapping expression patterns, MAB21-like proteins are expressed in distinct tissues. MAB21L1 is uniquely expressed in the lens vesicles and genital tubercle whereas MAB21L2 is found in the retinal epithelium, body wall and umbilical cord (178,179). Extensive mutational analysis proved MAB21-like protein's essential role in vertebrate embryogenesis. Zebrafish mutants have shown involvement of MAB21L2 in development of the midbrain-hindbrain boundary region as well as a proliferation defect within the retinal progenitor cell population (180,181). Whereas in *Xenopus* depletion of *Xmab21l2* leads to failure of gastrulation and impairment of notochord differentiation causing subsequent neural tube defects (174). *MAB21L2*-deficient mice have defects in eye and body wall formation causing abnormal extrusion of abdominal organs and resulting in death in mid-gestation (179,182). Another study demonstrated that MAB21L2 is essential for heart and liver development (183). In comparison *MAB21L1*-deficient mice show a relatively mild phenotype. Homozygous *MAB21L1* mutant mice evidenced defects in eye and preputial gland formation causing infertility (178). *MAB21L2*-/- as well as *MAB21L1*-/- defects are restricted to tissues with unique expression of one of the proteins, although transcripts are detected in other tissues. MAB21-like proteins compensate each other's function in shared expression domains, with MAB21L2 being more critical than MAB21L1 for embryonic development.

HOM-C/ Hox genes play a fundamental role in development. Two homeobox genes *egl-5* and *mab18* in *C. elegans*, homologous to *hox9-13* and *pax-6* in vertebrates, mediate cell fate decision events where *mab21* is required (164). In vertebrates the homeobox transcription factors Pax6, Rx and Mbx regulate MAB21L1 and MAB21L2 expression (180,184,185). Further experiments demonstrated that myeloid ELF1-like factor (MEF/ELF4) up regulates MAB21L1 on a transcriptional level whereas BMP2 down regulates MAB21L2 expression in osteoblasts (186,187). BMP2, 4 and 7 are members of the TGFβ signaling pathway and widely expressed in the gastrula ectoderm and mesoderm where they regulate dorsoventral axis formation as well as numerous key aspects in embryonic development. Regarding the molecular function of MAB21-like proteins, *Xenopus* MAB21L2 was reported to rescue ventralization of embryos by antagonizing BMP4 in its ability to ventralize embryonic mesoderm *in vivo*. Immunoprecipitation experiments *in vivo* with the BMP4 effector SMAD1, as well as *in vitro*, detected an interaction of MAB21L2 with SMAD1 and the SMAD1-SMAD4 complex (188) These results are in agreement with the epistatic interactions observed in nematodes (166,167).

A further MAB21 family member termed MAB21L3 was recently revealed to be required for the development of dorsal ectodermal and mesodermal fates in *Xenopus* (189). MAB21L1 and MAB21L2 share 40% and 44% sequence similarity with MAB21L3, respectively. Moreover, MAB21L3 acts downstream of the Notch pathway as key regulator in cell fate specifications of multiciliate cells (MCC) and ionocytes (190).

1.5.3 STRUCTURAL STUDIES ON MAB21L1

The first structural work on a MAB21 protein revealed that human MAB21L1 has an overall topology highly similar to its family member cGAS (FIG. 9). The bi-lobal structure with the typical α/β topology is composed of an N-terminal NTase domain with five-stranded, antiparallel and highly twisted β-sheets (β1-β9) flanked by two helices (α1/α4) and a C-terminal lobe formed by a conical four helix bundle (α8-α13). Both lobes are separated by a deep cleft and are hold together by the N-terminal 'spine' helix and a central linker region of two α-helices (α5, α6). The catalytic triad responsible for the Mg^{2+} ion coordination is only partially conserved in MAB21-like proteins (*hs*MAB21L1- E73 and E75). The third catalytic carboxylate residue, which normally coordinates the second Mg^{2+} ion and helps to polarize the acceptor hydroxyl group, is a catalytic inactive glutamine (*hs*MAB21L1- Q169) in the case of MAB21L1.

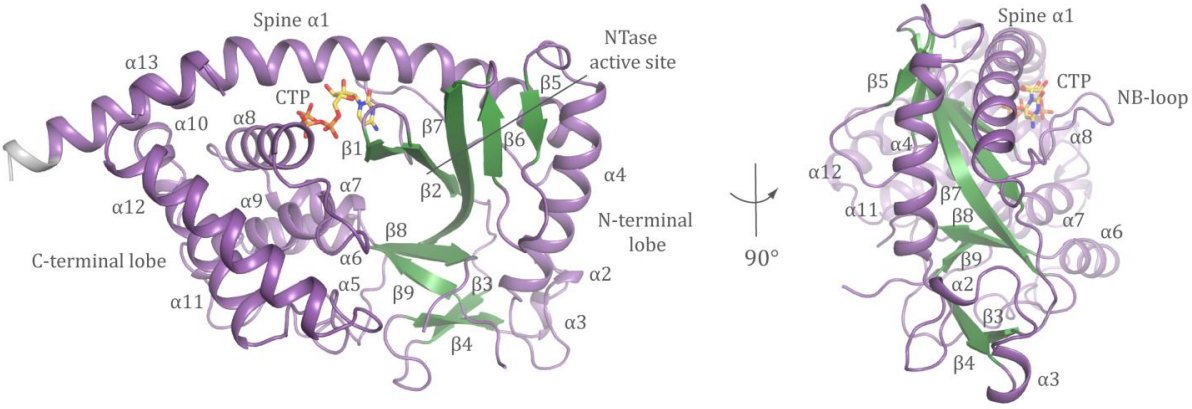


FIGURE 9: STRUCTURE OF MAB21L1 IN COMPLEX WITH CTP

Front and side views of the human MAB21L1 (purple, PDB: 5EOM) crystal structure in complex with CTP. The model is depicted as cartoon representation with annotated domains and secondary structure. CTP molecule is colored in yellow. Sequence in grey is reminiscent from the N-terminal tag.

In the crystal structure MAB21L1 assembles as a highly symmetric oligomer composed of two pentameric rings with back-to-back orientation (180° rotation) (FIG. 10). The MAB21L1 decamer has a D5-symmetry with one five-fold and five perpendicular two-fold non-crystallographic symmetry axes. The N-terminal lobe of each monomer is surface-accessible, while the C-terminal lobe is buried in the oligomer. The interaction interface of the pentamer is surprisingly small and distinct from that of other MAB21 domains known to dimerize (109,139,161,162). In contrast, the bulky stacking of the pentameric-rings in the crystal lattice is mediated between opposing lobes by the central linker region together with a strictly conserved region of MAB21L1 and MAB21L2. Despite different space groups and unit cell constants, this characteristic highly symmetric double-pentamer was observed in all MAB21L1 crystals raising the possibility that oligomerization might be a key function of MAB21L1. However, so far no experimental evidence of MAB21L1 oligomerization in solution is given.

Regarding the solvent accessible electrostatic surface of MAB21L1, there are several extensive positively charged regions. One notable positively-charged pocket stands out at the interface between the N-terminal 'spine' and the C-terminal lobe (ligand binding pocket). Similar, in cGAS there is a positively charged cleft in this region of the protein, which together with the Zn-thumb is responsible for DNA binding and recognition (106,135). In the MAB21L1 decamer another substantial positively charged surface around the five-fold symmetry axis is present. MAB21-like protein's positively charged surface correlates with their ability to bind oligonucleotides with preference for ssRNA (191). However, MAB21-like proteins bind generic oligonucleotides with considerable less affinity than cGAS suggesting that the ligand of MAB21-like proteins is more specific regarding sequence and/or structure.

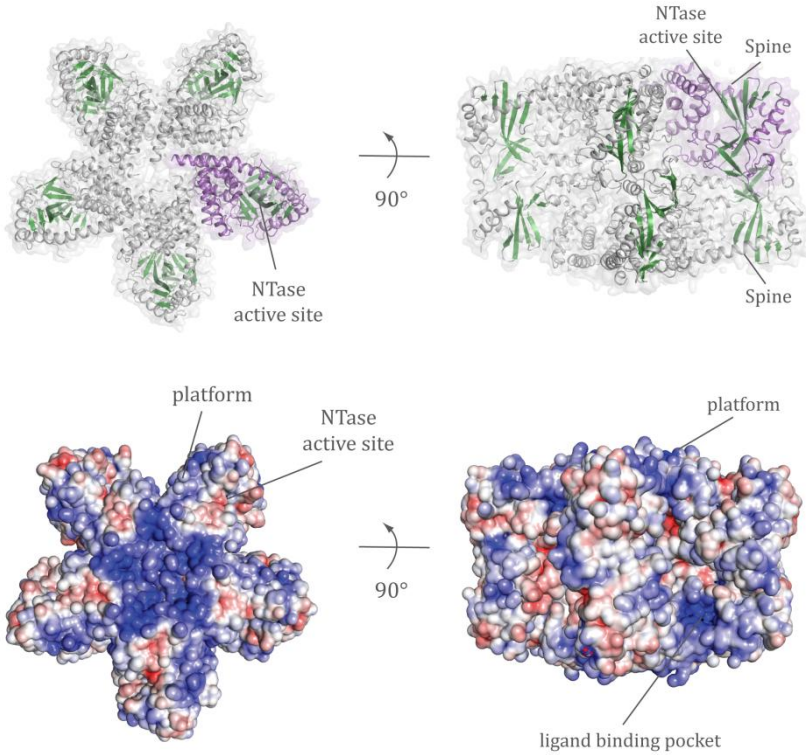


FIGURE 10: MAB21L1 OLIGOMERIZATION AND SOLVENT ACCESSIBLE ELECTROSTATIC SURFACE

Top panel: Front and side views of the MAB21L1 double-pentamer formation in the crystallographic asymmetric unit (PDB: 5EOM). One MAB21L1 monomer is colored in purple. Bottom panel: Corresponding surface accessible electrostatic surface representation (front and side views of the decamer) colored by charge (blue = 5kT/e to red = -5kT/e). The positively-charged ligand binding pocket and the positive patch correlating with the 'platform' in cGAS are indicated with an arrow.

Further experiments concerning a potential ligand/ substrate for the nucleotidyltransferase fold protein MAB21L1 evidenced an affinity for specific nucleotides. CTP as well as ATP were bound by MAB21L1, while other nucleotides, such as UTP and GTP, did not bind at all. A clear preference of MAB21L1 for ATP instead of ADP suggests that MAB21L1 ligand might contain a triphosphate. The structure of MAB21L1 bound to CTP exhibits a clear density for a single CTP moiety in each protomer. Interestingly, CTP was not located at the NTase active site. Instead, CTP bound in the positively charged pocket corresponding to the 'platform' in the

case of cGAS (FIG. 11). This 'platform' surface recognizes DNA in cGAS and RNA in OAS1 respectively, leading to major conformational changes required for activation of the enzyme (106,135,138). Nucleotide binding did not affect the overall conformation of MAB21L1, besides the nucleotide-binding loop which is better resolved in the CTP bound structure. This argues for an increased flexibility of this loop in apoMAB21L1 and CTP binding then traps MAB21L1's nucleotide-binding loop in an inactive conformation.

1.5.4 HIGH STRUCTURAL CONSERVATION BETWEEN MAB21L1 AND cGAS

The conserved MAB21 domain fold of MAB21L1 evidences high structural similarity to other nucleotidyltransferases, such as cGAS, OAS1 and MiD51/ MiD49 (136,162,163,192). Structural conservation between MAB21L1 and cGAS is striking despite of low sequence similarity. A sophisticated feature evolutionary inserted in cGAS, the Zn-thumb, is not present in MAB21L1. Active site residues located on the α-helical C-terminal lobe are highly conserved. These residues are responsible for 'donor' nucleotide coordination in other nucleotidyltransferases (FIG. 12). Y272 in *hs*MAB21L1, which is stacking to the ATP base, corresponds in porcine *ssc*GAS to Y413 (*hsc*GAS Y436). Further conserved residues essential for correctly positioning the 'donor' ATP in the active site of cGAS are: *ssc*GAS E360/*hsc*GAS E383 (MAB21L1 E238), which holds the ribose in place, *ssc*GAS K416/*hsc*GAS K439 (MAB21L1 K275) and *ssc*GAS K391/ *hsc*GAS K414 (MAB21L1 K255) responsible for coordination of the negative charge of the phosphate groups.

Striking differences are found in the N-terminal lobe, which coordinates the Mg^{2+} ions and mostly contacts with the 'acceptor' GTP molecule. While the adenine base is not involved in any intermolecular hydrogen bonds, the guanine base is in hydrogen bond distance with its Watson-crick and Hoogsteen edges from residues *ssc*GAS T186, S57, S355 and R353 (*hsc*GAS T211, S378, R376 and S380) respectively. Notably, these residues, which confer cGAS specificity for the guanine base, are missing in MAB21L1. T186 cannot be assigned, since the nucleotide binding loop is in an inactive conformation in MAB21L1 and the other residues are A235, V231 and Q233, in MAB21L1 respectively. As mentioned before, the catalytic triad coordinating the two Mg^{2+} ions that hold the ATP triphosphate in place is only partially conserved in MAB21 proteins. In *ssc*GAS the first Mg^{2+} ion is coordinated by E200

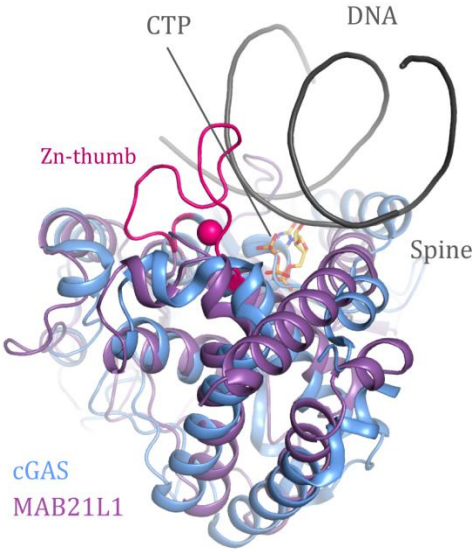


FIGURE 11: POTENTIAL LIGAND BNDING POCKET IN MAB21L1 CORRELATES WITH 'PLATFORM' IN cGAS

Superposition of MAB21L1 (purple, PDB: 5EOM) in complex with CTP (yellow) and cGAS (blue, PDB: 4KB6) in complex with dsDNA (grey). Zn-thumb is depicted in pink.

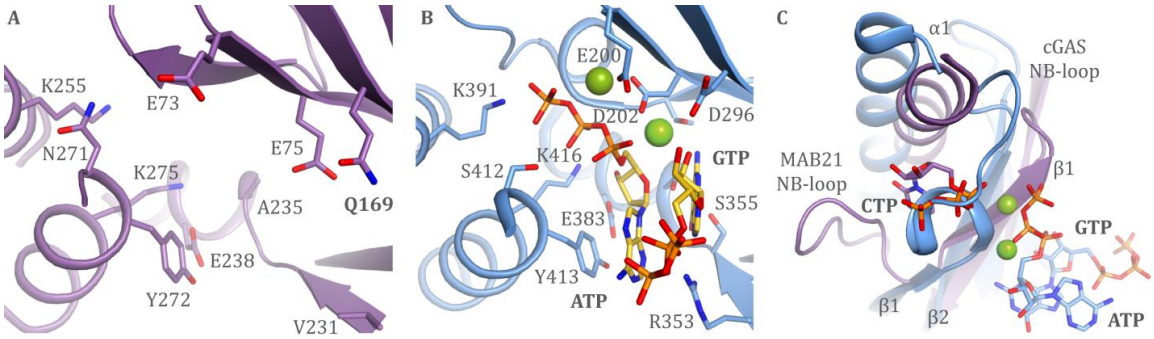


FIGURE 12: ACTIVE SITE OF MAB21L1 IN COMPARISON WITH STRUCTURALLY CONSERVED cGAS

A. Close up view of the NTase active site of human MAB21L1 (purple, PDB: 5EOM) in an inactive conformation with annotated residues implicated in catalysis and conserved to cGAS. B. Close up view of the NTase active site of porcine cGAS-DNA-ATP-GTP (blue, PDB: 4KB6) in the same orientation. Q200 and N202 are mutated in figure to E200 and D202, respectively, and a second Mg^{2+} is added to the catalytic site for reason of clarity. C. Close-up view of distinct nucleotide binding sites in the cGAS-DNA-ATP-GTP and MAB21L1-CTP complexes.

and D202 (*hs*cGAS E225 and E227, MAB21L1 E73 and E75) and D296 facilitates the coordination of the second Mg^{2+} ion. MAB21L1 has the third catalytic carboxylate residue changed to an inactive glutamine Q169.

The MAB21L1 structure bound to CTP represents an inactive conformation. In structures of inactive cGAS or OAS1 the loop connecting β1 and β2 is not resolved (106,135,192). In MAB21L1 this loop is resolved, due to CTP binding. cGAS and OAS undergo large conformational changes upon activation and especially part of the nucleotide-binding loop (β1-β2 hairpin) becomes a α-helical turn (146). CTP binding to MAB21L1 prevents these rearrangements and access to the catalytic pocket with residues such as I69 hindering coordination of the catalytic Mg^{2+} ions. It is still unclear, if MAB21L1 features nucleotidyltransferase activity, despite the lack of a complete catalytic triad and being in an inactive conformation.

1.5.5 PATIENT MUTATIONS IN MAB21L2 CAUSE EYE MALFORMATIONS

Recently, three independent whole exome sequencing projects identified four different MAB21L2 missense mutations in patients with bilateral ocular coloboma (191). Eight individuals from five unrelated families all evidenced mutations altering only one gene (MAB21L2) out of a previously compiled list of 38 candidate genes for eye malformations. Three mutational events altered Arg51 to cysteine or histidine (R51C/H) leading to bilateral anophtalmia, intelectual disability and rhizomelic skeletal dysplasia (193). Another further mutation of Glu49 mutated to lysine (E49K) was located within two residues of Arg51 and in a different region of the protein; Arg247 is mutated to glutamine (R247Q). Both mutations were associated with retinal coloboma. These mutations support a strong involvement of MAB21L2 in human ocular development. First attempts to clarify MAB21L2's role in this drastic phenotypes based on alignments of MAB21L2 with cGAS and OAS1 (191). In

agreement with the location of the mutations in MAB21L2 close to the RNA binding cleft in OAS1, all mutations were able to disrupt the affinity of MAB21L2 for ssRNA (191). In addition, Rainger et al. generated GFP fused MAB21L2 mutants (R51C/H, E49K, R247Q) that proved to be more stable than the wild type protein (191). In contrast, subsequent experiments in zebrafish evidenced a decreased stability of the mutant R51G compared to wild type MAB21L2 (194).

Since only mutation R247Q located on the C-terminal lobe is structurally conserved in cGAS, the crystal structure of MAB21L1 brought further insights into the intriguing question whether the mutations have stabilizing or destabilizing effects. All amino acids mutated in MAB21L2 are conserved in MAB21L1. Instead of being located at the surface as previously proposed, the mutated residues are involved in structure stabilization (FIG. 13). From the MAB21L1 structure it is visible that R247, which is structurally conserved in cGAS, forms a salt bridge with E288 and that E49 together with R51 participate in a network of salt bridges with E115.

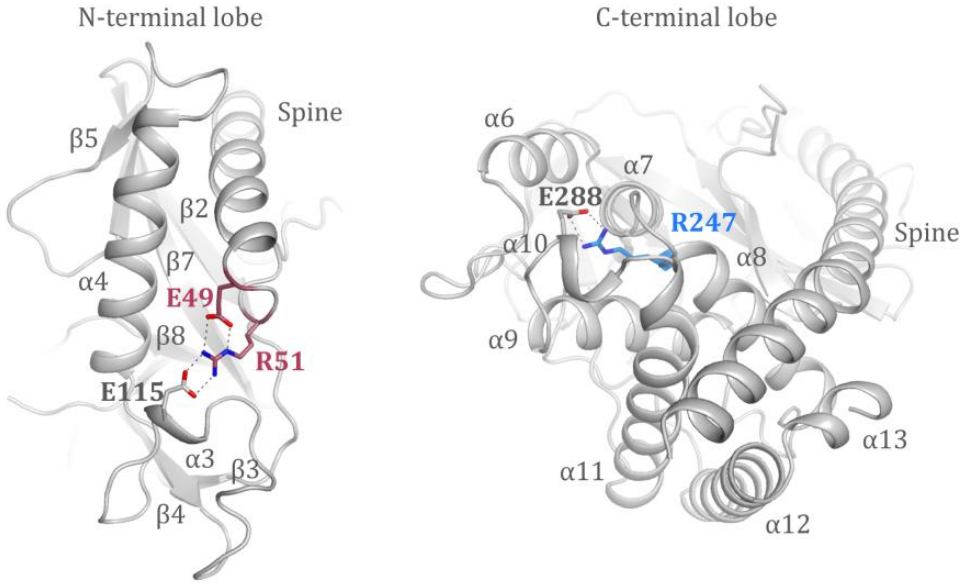


FIGURE 13: PATIENT MAB21L2 MUTATIONS MAPPED ON MAB21L1 STRUCTURE

Amino acids of MAB21L2 mutated in patients with eye malformations are depicted in red (E49K, R51C/H) on the N-terminal lobe and in blue (R247Q) on the C-terminal lobe of the MAB21L1 structure.

All mutational events disrupt the aforementioned integral structural interactions. Thermal shift assays of MAB21L2 wild type, R51C and R247Q demonstrated that mutations lead to a decrease in stability of the protein. In particular, R247Q resulted in even more drastic destabilization compared to R51C. R247 seems to be more fundamental for structural stability of the MAB21 domain fold *per se*. The highly conserved R51 and E49 residues are responsible for correctly positioning the loop β4-α4 and allowing the two α-turns α2 and α3 (FIG. 14). The high degree of conservation of this loop suggests that correct folding is of major importance for the stability and proper function of MAB21-like proteins.

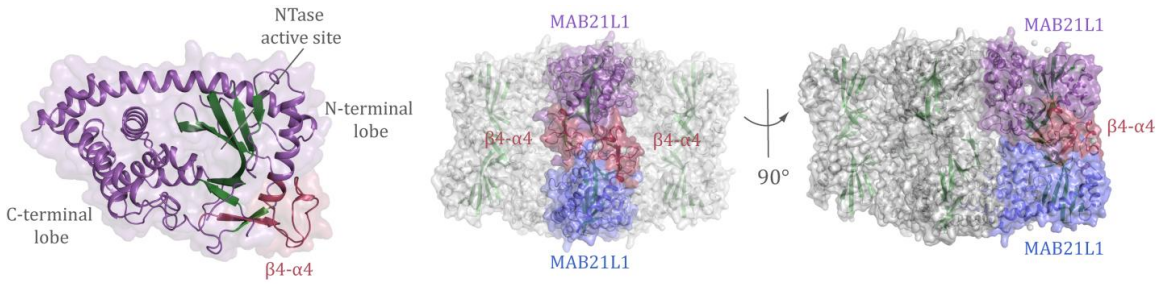


FIGURE 14: PATIENT MUTATIONS IN MAB21L2 AFFECT THE FOLD OF CONSERVED LOOP

Left panel: MAB21L1(purple, PDB: 5EOM) monomer with the β4-α4 loop marked in red. Center and right panel: MAB21L1 decamer with one MAB21L1 protomer in purple and the neighboring MAB21L1 protomer from the adjacent pentamer in blue. β4-α4 loops of both MAB21L1 interact and are depicted in red.

1.6 MECHANISM OF CYTOSOLIC RNA SENSING BY RIG-I

1.6.1 STRUCTURAL MECHANISM OF RNA SENSING AND RLR ACTIVATION

RIG-I surveys the cytoplasm in an auto-inhibited state where the helicase is in an open conformation (49,58). The CTD is flexibly linked to the helicase having no major interactions with the rest of the protein, while the CARDs are sequestered by Hel2i through extensive contacts between the top end of CARD2 (α23, α24) and α25 of Hel2i (54). In the inactive state CARD1 is mainly surface accessible. In contrast, the proximity of CARD2 to the Hel1 domain as well as the Hel1-linker and the Hel2i interface sterically hinder interactions with other macromolecules (MAVS-CARD, polyubiquitin) and poly-ubiquitination required for downstream signaling. Indeed, disruption of the CARD2-Hel2i interface led to constitutively active RIG-I and stimulation of the IFN β promoter in human Huh7.5 cells (54). Moreover, the tandem CARDs form a rigid functional unit which is essential as a whole for downstream signaling events (69). RIG-I's auto-inhibited state is supported via masking of the RNA binding site on the Hel2i domain by the CARDs. This contributes to inhibition of dsRNA binding and dsRNA-dependent ATPase activity of RIG-I. ATPase activity is lacking in inactive RIG-I due to the stabilized open conformation where the helicase motifs are misaligned (58).

The first step in RIG-I activation is the capturing of the RNA by the flexible and surface exposed RD (CTD), which selectively binds 5´ppp dsRNA (195,196). Binding of the RNA to the RD increases the local concentration of dsRNA near the helicase and favors therefore the co-operative binding of ATP and dsRNA. Following major rearrangements lead to a closed conformation of RIG-I where the helicase-RD organizes into a ring around the dsRNA capping the blunt end 5´ppp dsRNA and simultaneously liberating the CARDs (FIG. 15) (54,56,57). The crystal structure of RIG-I helicase-RD in complex with ADP•BeF_3, with BeF_3 mimicking the γ-phosphate of ATP, shows how all four domains (Hel1, Hel2, Hel2i and RD) move to each

other upon binding dsRNA and ATP (56). Conformational changes include rotation of Hel1 by 49° from Hel2 as a reference and rotation of Hel2i by 24° (54). The dsRNA maintains its helical A-form conformation in a highly basic channel of the helicase-RD containing previously uncharacterized helicase motifs for RNA recognition. The 3´ strand is exposed and mainly interacts with the helicase, whereas the 5´end harboring the tri-phosphate is capped by the helicase-RD. Interactions between RIG-I's helicase-RD and dsRNA are mainly sequence-independent involving the ribose-phosphate backbone of both strands. Residues known to interact with dsRNA in previous crystal structures of the isolated RD are identical in the helicase-RD structure bound to dsRNA (195,197,198). However, the trajectory of the dsRNA in the isolated RD structures differs from the one in the helicase-RD structure where the dsRNA bound to the isolated RD would clash with Hel2i. Hence, the dsRNA in helicase-RD is rotated and new contacts with the RD are introduced. Furthermore, the critical RNA binding loop (849-857) containing F855, which stacks on the blunt-end bases, changes its conformation upon dsRNA binding allowing capping of the blunt-end dsRNA (56). In the helicase-RD structures, a two α-helical V-shaped linker that extends into a proline-rich loop makes the connection from Hel2 to the RD. RIG-I's activity was shown to be regulated through phosphorylation of T770 by Casein kinase II (199). This residue is located on the kink of this V-shaped linker (pincer) suggesting a mechanistic role for the linker such as a hinge for RD positioning (54,56).

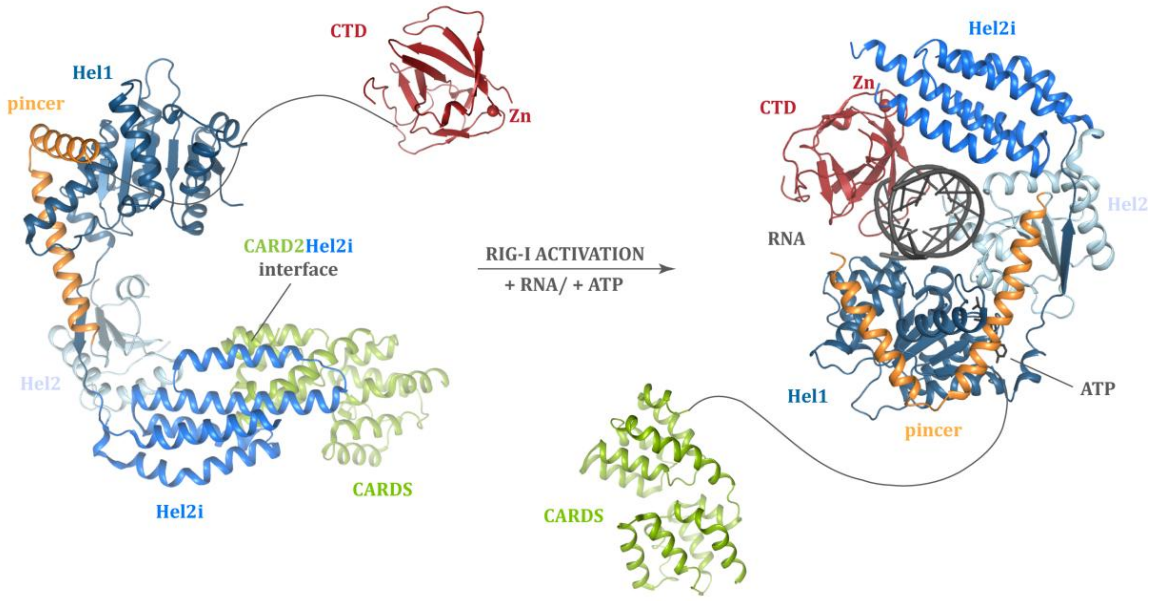


FIGURE 15: MODEL FOR THE ACTIVATION OF RIG-I

Inactive RIG-I (PDB: 4A2Q) has an open conformation where the CARDS (green) are sequestered by the Hel2i (blue). The pincer (orange) holds RIG-I in the inactive conformation where the CTD (red) is flexibly linked to the helicase (blue). Upon 5´ triphosphate dsRNA and ATP binding RIG-I (PDB: 5E3H) adopts a closed conformation with the helicase (blue) forming a ring around the dsRNA (grey). The CTD (red) displaces the CARDS (green) from the Hel2i interface. The CARDS (green) are then accessible for downstream signaling.

Regarding the helicase core domains, characteristic motifs interact with the dsRNA. Hel1 domain motifs Ia, Ib and Ic interact with the 3´ strand sugar-phosphate backbone, whereas

motif IIa and motif II are in contact with the 5´ strand. In the Hel2 domain, motifs IV, IVa, IVb and V interact with 3´ strand from C13 to C10 and the 5´ strand is bound by motif Vc (54,56). These newly characterized RNA recognizing motifs may constitute a general feature for dsRNA or dsDNA binding helicases, since motif IIa was identified to be involved in dsDNA binding by a Rad54 homologue belonging to the Swi/ Snf2 family of helicases (200). Motifs Q, I, II, III, Va and VI are generally involved in ATP binding and hydrolysis. In the RIG-I helicase-RD structure ADP•BeF_3 is bound in the interface between Hel1 and Hel2 and nucleotide coordination is achieved by residues in the responsible helicase motifs. The adenine is specifically recognized by motif Q, phosphates together with BeF_3 are bound by motif I, the Mg^{2+} ion is coordinated by motif II (D372 and E373) and residues in motifs Va and VI generally contact with the ribose and phosphates of the ATP analog (54,56).

Hydrogen/deuterium exchange coupled with mass spectrometry (HDX-MS) experiments evidenced that in RIG-I the resting auto-inhibited state is maintained by CARD2-Hel2i interactions, while in MDA5 the analogous intra-molecular interactions are lacking (201). This results in an extended conformation of MDA5 where the CARDS are exposed. In addition, human MDA5 has a significantly longer linker between CARD2 and the helicase compared to RIG-I (105 and 55 residues respectively). Hence, RIG-I and MDA5 have different resting states and thus mechanisms of activation.

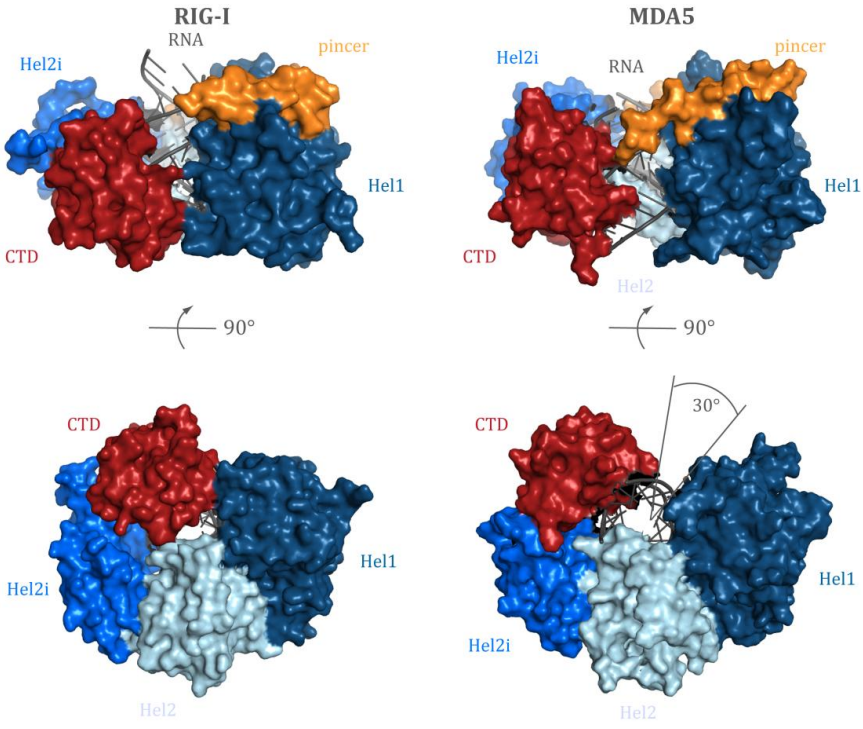


FIGURE 16: COMPARISON OF RIG-I AND MDA5 RNA BINDING MODES

Side and bottom views of RIG-I (left, PDB: 5E3H) and MDA5 (right,PDB: 4GL2) bound to dsRNA (grey). RIG-I forms an asymmetrically closed O-ring around the dsRNA. In the dsRNA end capping mode of RIG-I, the CTD (red) of RIG-I is tilted towards the dsRNA end. Instead, MDA5 forms an open ring structure (C-shaped, 30° opening between CTD and Hel1) around the dsRNA stem orienting the CTD (red) parallel to the dsRNA axis.

Concerning the dsRNA recognition, major differences in the orientation of the CTD enable both binding to dsRNA stem by MDA5 and capping of dsRNA ends by RIG-I (FIG. 16) (54,56,57,202). RIG-I forms an asymmetrically closed O-ring around the dsRNA where the bottom tip of the CTD is tilted towards the dsRNA end close to Hel1. In contrast, the CTD of MDA5 is placed parallel to the axis of dsRNA forming a C-shaped ring around the dsRNA stem (203). The Hel2i domain provides a docking site for the CTD of MDA5 to be relocated near the dsRNA stem. In the case of RIG-I, the Hel2i interface has an additional protrusion in helix α12 that is able to push the CTD away from Hel2i preventing the same orientation of the CTD for RIG-I (202). Another difference represents the 5´ppp and blunt end recognizing RNA binding loop (849-857) in RIG-I, which is disordered in MDA5 leaving dsRNA stem recognition to the flat surface of the CTD. A loop located in the Hel2 domain plays an important role in coupling RNA binding with ATP hydrolysis and subsequent signaling. In MDA5 the Hel2 loop inserts into the major groove of the dsRNA stem, whereas in RIG-I it binds the dsRNA end. Besides the distant location of the Hel2 loop from the ATP-binding site, deletion of this loop abolishes dsRNA-dependent ATP hydrolysis (202).

1.6.2 ATP HYDROLYSIS BY RIG-I PREVENTS RECOGNITION OF SELF-RNA

RLRs are classified as RNA helicases belonging to the structurally conserved helicase superfamily 2 (often referred to as DExH/D or DExD/H proteins) (204). The term 'helicase' was first proposed to describe ATP-dependent duplex DNA unwinding activity (205). In particular, SF2 members are characterized by nucleic acid binding and simultaneous ATP hydrolysis coupled to large conformational changes triggering downstream events (206). Helicase conformational changes generally include structural remodeling of bound nucleic acid, protein-nucleic acid complexes or translocation of the helicase along the bound nucleic acid. However, not all helicases or members of the DExD/H helicase family share this translocation-dependent unwinding nor unwinding activity alone (206). Single-molecule analysis revealed that RIG-I has ATP-dependent translocation activity reflected in transient rapid on/off interactions with the duplex RNA stem (207). Further, in the absence of 5´-triphosphate, translocation was drastically suppressed by the CARDs. RIG-I as well as MDA5 seem to lack duplex unwinding activity, besides a report showing RIG-I to unwind short duplex 3´ overhang RNAs in an ATP-dependent manner (208).

RIG-I mediated signaling was shown to systematically increase with dsRNA length at a constant molar amount (209). Involvement of the dsRNA stem in RIG-I activation together with the ATP-dependent translocation activity suggest that the helicase domain of RIG-I might be capable of low affinity internal binding to dsRNA devoid of a 5´-triphosphate end. In fact, dsRNA termini are bound by RIG-I as a monomer in the absence of ATP and upon ATP hydrolysis, RIG-I was demonstrated to assemble into filamentous oligomers that propagate from the dsRNA end to the interior (210,211). Importantly, dsRNA ends and ATP hydrolysis are continuously required for filament formation, explaining how previous studies of RIG-I without ATP led to the misleading notion of monomeric RIG-I bound to dsRNA (210). Crystal

structures of RIG-I capping the dsRNA end are consistent with dsRNA translocation without unwinding and represent the conformation of RIG-I prior to translocation and filament formation (54,56,57). Unfortunately, no crystal structures depicting the conformation of RIG-I bound internally to long dsRNA are available. End-capping structures of RIG-I are incompatible with filament architecture suggesting that RIG-I dsRNA stem-binding requires a different orientation of the CTD, which might be similar to that of MDA5 for dsRNA stem recognition (202,210). MDA5 was shown to distinguish between dsRNA and other nucleic acids through cooperative assembly into a filamentous oligomer along the length of the dsRNA (202,212). MDA5 stacks along the dsRNA in a head-to-tail arrangement where the Hel2i domain of one protomer binds the pincer domain of the next protomer. In contrast, RIG-I filaments evidence no protein-protein interactions (202,210). Further, the impact of ATP hydrolysis on RIG-I and MDA5 oligomerization states is remarkably different. In the case of MDA5, filament formation is ATP-independent, while ATP hydrolysis dynamically regulates MDA5's disassembly process in a dsRNA length-dependent manner. The ATP-driven length dependent dissociation enables MDA5 to measure the dsRNA length and renders short dsRNA kinetically unstable for signaling via MDA5 (202,212). Instead, RIG-I requires ATP hydrolysis for translocation and associated filament formation (210,211).

ATP hydrolysis is associated with the conformational switch required for signaling of RIG-I as well as MDA5 (69). In several studies ATPase defective and gain-of-function alleles of RIG-I and MDA5 result in either improved viral clearance or autoimmune diseases including Aicardi-Goutières and Singleton-Merten syndromes (SMS) (213-216). While SMS mutations resulted in constitutive RIG-I signaling described as gain-of-function, previous studies dealing with mutations in helicase motif I rendered inactive RIG-I, different types of mutations in motif II resulted in either gain or loss of function and finally mutations in motif III, which significantly disrupt ATPase function had no impact on RIG-I signaling (FIG. 17) (31,48,58,217). Mutation K270I in motif I, which abrogates ATP binding, is unable to signal when challenged with longer RNAs (217-219). Short 10bp RNA did not abolish signaling, suggesting that Walker A mutants are only detrimental to signaling on longer RNA (219). The structure-derived E373Q mutation in motif II, however, resembles the phenotype of SMS mutations. E373 in motif II is required for proper Mg^{2+} ion coordination to stabilize ATP. Hence, RIG-I E373Q mutant is trapped in an ATP-bound state leading to constitutive signaling in response to self-RNA (218). ATPase activity helps both helicases to discriminate between self and non-self dsRNA by fine-tuning release of the CARDs and thus the hereby associated downstream signaling. Low ATP hydrolysis in RIG-I or MDA5 mutants correlates with constitutive signaling. This effect is reverted, when mutations that disrupt RNA binding to the helicase are inserted (217). RIG-I mutations T347A and V699A which disrupt the interaction of the RNA with either the Hel1 or Hel2 domain respectively, were designed to test if E373Q signaling in non-infected cells is due to interactions with self-RNA. T347A mutation disrupted signaling in either infected or non-infected cells, whereas V699A slightly increased the signaling (218). The latter can be related to a putative reduction in translocation ability instead of the desired disruption of RNA binding. In order to dissect, if binding to endogenous RNA by RIG-I (E373Q) is dependent on RNA harboring a 5´-triphosphate or 5´-diphosphate, K888T was mutated (218). 5´triphosphate binding in the RD is mediated by residue K888

(56). RIG-I E373Q/ K888T double mutant signals constitutively in non-infected cells, indicating that signaling upon binding to self-RNA is independent from 5´triphosphates (218). Another autoimmune disease related mutation C268F located on motif I, is constitutively active (218). Intriguingly, previous mutations in motif I (K270I) abolished RIG-I signaling. The non-ATP binding C268F mutation, behaves as a motif II mutation (E373Q) and therefore phenocopys a mutation that prevents ATP hydrolysis.

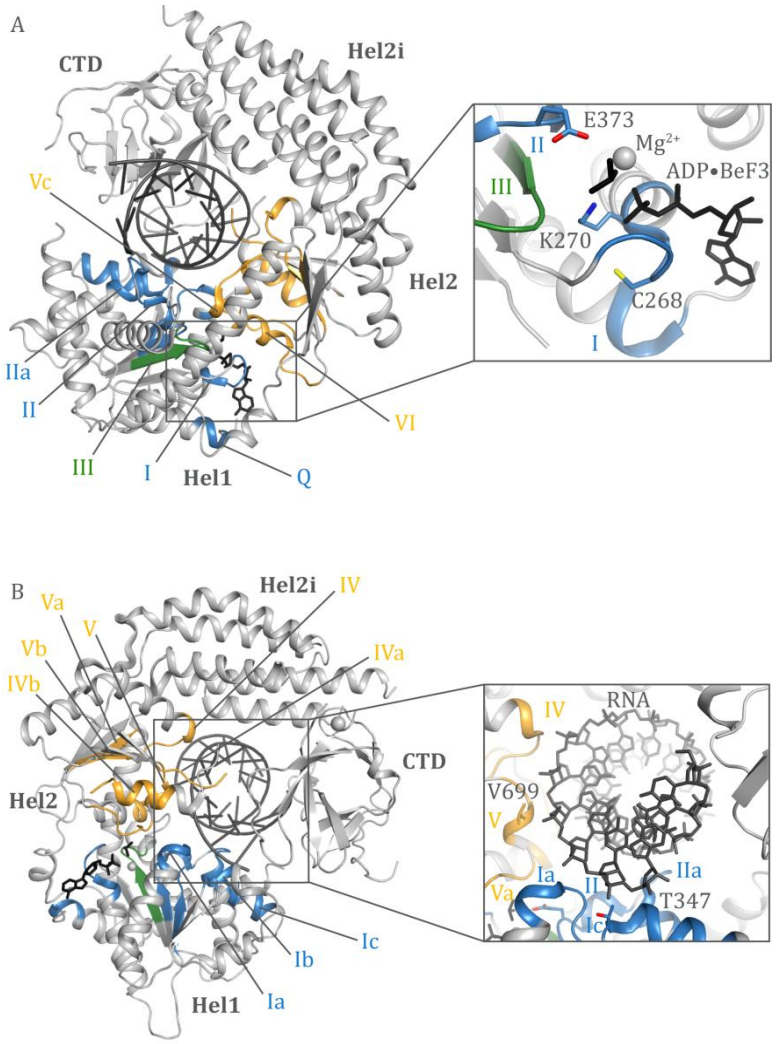


FIGURE 17: RIG-I HELICASE MOTIFS

A and B. Structure RIG-I bound to dsRNA (grey) and ADP•BeF_3 (black, PDB: 5E3H) with helicase motifs responsible for RNA and ATP binding denoted in blue (Hel1, motifs Q, I, Ia, Ib, Ic, II, IIa), green (motif III) and in yellow (Hel2, motifs IV, Iva, IVb, V, Va, Vb, Vc and VI). A. Inset shows a close up of the ATP binding site of RIG-I. Residues mutated in this work are denoted. B. Inset shows a close up of the dsRNA binding site. Residues mutated in this work are denoted.

The nature of the endogenous RNA ligand for RIG-I E373Q was characterized from co-immunoprecipitation experiments in non-infected cells (218). E373Q mutants co-purified with a three-fold increased amount of RNA compared to wt RIG-I. This RNA was not immunostimulatory in a wt RIG-I background and most of it was uncovered to be 28S ribosomal RNA. Cryo-EM analysis of RIG-I E373Q mutant in complex with the ribosome revealed an extra density for RIG-I E373Q located at the expansion segment ES-7L (218). Ribosomal expansion segments have G:C rich, base-paired RNA stretches that form large tentacle-like hairpin structures with considerable double-stranded content, therefore providing an optimal dsRNA ligand for RIG-I stabilized in the ATP-bound state (220).

ATP binding to the helicase domain contributes to the overall affinity of RIG-I to dsRNA in addition to the RD, which dominates the interactions with dsRNA. This notion is supported by several structures of RIG-I in complex with ATP analogs (54,56). These structures of RIG-I depict a progressive closure of the helicase domain upon ATP binding which is responsible for triggering the release of the tandem CARDs. Moreover, the CARDs play a major role in mediating the dissociation from RNA by linking it to ATP binding to the helicase. In agreement with this, ΔCARDs constructs showed no enhanced dissociation from RNA in the presence of ATP (219). ATP addition shows no effect on the affinity to 5´-triphosphorylated dsRNA, suggesting that upon ATP binding and subsequent hydrolysis only RIG-I dissociation from unphosphorylated RNA ends is promoted (218). Therefore inhibition or slow-down of the ATPase activity of RIG-I leads to prolonged interaction periods with endogenous dsRNA ligands devoid of 5´-triphophate epitopes and to the corresponding constitutive downstream signaling.

2 Publications

2.1 STRUCTURAL MECHANISM OF CYTOLSOLIC DNA SENSING BY cGAS

Civril, F.*, Deimling, T.*, de Oliveira Mann, C. C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.-P. (2013) Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332-337

*: equal contribution

This publication reports the crystal structure of porcine cGAS (MAB21-domain) alone and in complex with a 14-mer dsDNA, ATP and GTP along with functional studies *in vitro* and in human cells. Recent studies identified the cyclic GMP-AMP synthase (cGAS) as the major cytosolic DNA sensor. In the presence of dsDNA, cGAS binds ATP and GTP and subsequently catalyses the production of a novel second messenger cGAMP bearing a noncanonical 2´-5´ and a 3´-5´ mixed phosphodiester linkage. Upon cGAMP binding, STING dimerizes and stimulates a robust IFN response to cytosolic DNA. Our results explain the broad sequence-independent sensing specificity of cGAS along with the activation mechanism based on a structural switch induced by DNA binding. Further, the crystal structure of a transferase-trapping active site mutant together with the substrate nucleotides readily defines how cGAS first catalyses the synthesis of a linear dinucleotide intermediate. Taken as a whole these results mechanistically unify dsRNA and dsDNA recognition in innate immunity by cGAS and OAS1. A new unique class of conserved of 2´-5´-linked second messengers synthesizing enzymes is revealed with similar structural fold and activation mechanisms. cGAS contains a unique Zn-thumb that serves as molecular 'ruler' and specific residues involved in minor-groove interactions enabling it to recognize the backbone of B-form dsDNA. A-form dsRNA is not able to trigger the conformational changes required for cGAS activation and for proper orientation of the active site geometry in order to catalyse cGAMP.

Author contribution

The author of this thesis cloned, expressed and purified the porcine and human cGAS constructs and mutants for crystallization, *in vitro* and cell-based assays. She performed electro-mobility shift assays (EMSA) in order to characterize binding affinities for different RNA/DNA species of wild type cGAS, STING and the effects of several cGAS mutations.

ARTICLE

doi:10.1038/nature12305

Structural mechanism of cytosolic DNA sensing by cGAS

Filiz Civril[1]*, Tobias Deimling[1]*, Carina C. de Oliveira Mann[1], Andrea Ablasser[2], Manuela Moldt[1], Gregor Witte[1], Veit Hornung[2] & Karl-Peter Hopfner[1,3]

Cytosolic DNA arising from intracellular bacterial or viral infections is a powerful pathogen-associated molecular pattern (PAMP) that leads to innate immune host defence by the production of type I interferon and inflammatory cytokines. Recognition of cytosolic DNA by the recently discovered cyclic-GMP-AMP (cGAMP) synthase (cGAS) induces the production of cGAMP to activate the stimulator of interferon genes (STING). Here we report the crystal structure of cGAS alone and in complex with DNA, ATP and GTP along with functional studies. Our results explain the broad DNA sensing specificity of cGAS, show how cGAS catalyses dinucleotide formation and indicate activation by a DNA-induced structural switch. cGAS possesses a remarkable structural similarity to the antiviral cytosolic double-stranded RNA sensor 2′–5′oligoadenylate synthase (OAS1), but contains a unique zinc thumb that recognizes B-form double-stranded DNA. Our results mechanistically unify dsRNA and dsDNA innate immune sensing by OAS1 and cGAS nucleotidyl transferases.

Recognition of pathogen- or danger-associated molecular patterns (PAMPs or DAMPs) is crucial for host defence. Innate immunity ensures this recognition through germline-encoded pattern recognition receptors (PRRs) and triggers signalling cascades that result in production of proinflammatory cytokines and type I interferons (IFN-α and IFN-β)[1,2]. Cytosolic DNA arising from intracellular bacteria or viral infections is a powerful PAMP and is also implicated as a DAMP in autoimmune diseases[1,3,4]. Over the past years, a variety of PRRs for cytosolic DNA have been reported: DNA-dependent activator of IFN-regulatory factors (DAI, also known as ZBP1)[5], absent in melanoma 2 (AIM2)[6–8], RNA polymerase III[9,10], leucine-rich repeat (in Flightless I) interacting protein-1 (LRRFIP1)[11], DExD/H box helicases (DDX41, DHX9 and DHX36)[12,13] and IFN-inducible protein IFI16[14]. However, these PRRs are either cell-type- or DNA-sequence-specific, are possible accessory factors (DExD/H proteins), or trigger different pathways such as caspase-1 activation (AIM2) or a β-catenin-dependent signalling pathway (LRRFIP1)[15].

Although the DNA sensor for type I IFN production with broad specificity and cell distribution was not identified until recently, it was known that IRF3 and NFκB activation in response to DNA requires STING (stimulator of interferon genes, encoded by gene *TMEM173* the protein is also known as MITA, MPYS or ERIS), a transmembrane protein that is resident on the endoplasmic reticulum[16–18]. STING colocalizes with DNA *in vivo* but binds DNA only with low affinity *in vitro*[19], suggesting the presence of an additional sensor. Furthermore, STING is a direct PRR for cyclic dinucleotides such as c-di-AMP and c-di-GMP[20], which are signalling molecules in prokaryotes and trigger IFN in response to, for example, intracellular bacteria[21,22].

Recent results identified human c-GMP-AMP (cGAMP) synthase (cGAS, also known as C6ORF150 and male abnormal 21 domain containing 1 (MB21D1)) as a broad-specificity cytosolic DNA sensor[23]. In the presence of DNA cGAS produces cGAMP, which is an endogenous second messenger that activates STING[18], explaining how STING can stimulate IFN in response to both cyclic dinucleotides and DNA. To reveal the mechanism of DNA-stimulated cGAMP synthesis, we determined the crystal structure of porcine cGAS^{Mab21} (residues 135–497, comprising the highly conserved, DNA-stimulated nucleotidyl transferase (NTase) domain) with and without a 14-mer dsDNA ligand and nucleotide substrates, along with functional studies *in vitro* and in living cells.

Crystal structure of cGAS^{Mab21}

cGAS is a 60 kDa protein composed of an unstructured, not well conserved amino-terminal stretch of approximately 130–150 residues followed by a highly conserved Mab21 domain that belongs to the nucleotidyl transferase (NTase) superfamily[24]. To overproduce and crystallize cGAS, it was necessary to genetically remove the unstructured N-terminal tail. The resulting cGAS^{Mab21} used in this study (residues 155/161–522 for human cGAS and residues 135–497 for porcine cGAS) possesses DNA-dependent dinucleotide synthesis activity in the presence of a 50-mer dsDNA that induces IFN in THP1 cells (Fig. 1a and Supplementary Fig. 1a, b). Whereas cGAS also produces cGAMP in the presence of a 40-mer dsDNA, no activity was observed when we omitted either GTP or ATP from the reaction mixture or substituted dsDNA with single-stranded DNA (Supplementary Fig. 1a).

We determined the crystal structure of porcine cGAS^{Mab21} by single-wavelength anomalous dispersion to 2.5 Å resolution using a selenomethionine derivative. After density modification, we could build an initial model, which was completed and refined against the 2.0 Å resolution native data, resulting in good *R*-factors and stereochemistry (Supplementary Fig. 1c and Supplementary Table 1).

The Mab21 domain of cGAS comprises two lobes, separated by a deep cleft (Fig. 1b). Lobe 1 possesses the NTase fold with a two-leaved highly twisted β-sheet (β1–β8) that is flanked on the outside by two long α-helices (αA and αB). At the inner side, lining the cleft, β1 and β6 harbour the signature catalytic site residues (E200, D202, D296) of the NTase superfamily that coordinate the catalytic Mg^{2+} ions and nucleotides. Lobe 2 is a bundle of four α-helices (αE–αH), connected

[1]Department of Biochemistry and Gene Center, Ludwig-Maximilians-University, 81377 Munich, Germany. [2]Institute for Clinical Chemistry & Clinical Pharmacology, Unit for Clinical Biochemistry, University Hospital, University of Bonn, 53127 Bonn, Germany. [3]Center for Integrated Protein Sciences, 81377 Munich, Germany.
*These authors contributed equally to this work.

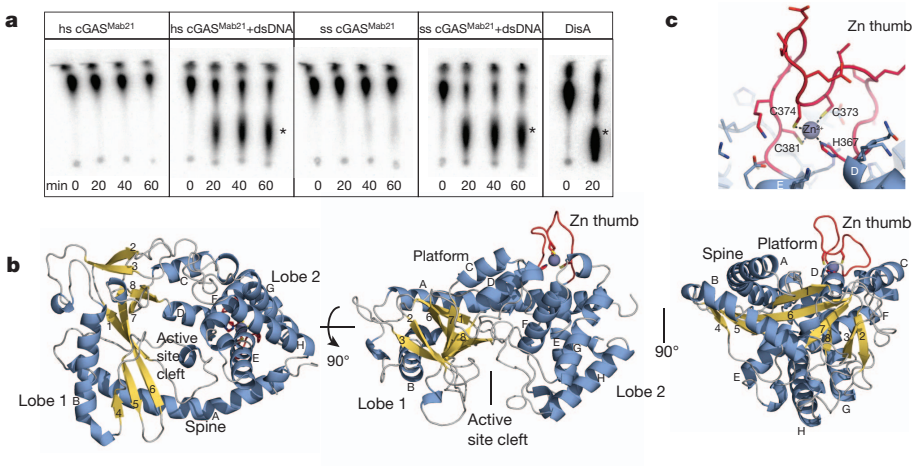


Figure 1 | Crystal Structure of cGAS^{Mab21}. **a**, Activity assays of human and porcine cGAS^{Mab21} alone or in presence of dsDNA. *Bacillus subtilis* DisA, a c-di-AMP synthase is used as positive control. The dinucleotide products are indicated with asterisks. **b**, Side and top views of cGAS^{Mab21}. The model is shown as ribbon representation with annotated domains and secondary structure (blue α-helices, yellow β-strands). **c**, Close-up view of the 'zinc thumb'.

to lobe 1 by a long 'spine' (αA), two linker helices (αC, αD) and by a long active site loop connecting αA and β1.

The molecular surface opposite the active site is a fairly flat, slightly concave 'platform', formed predominantly by αA, αC, αD and the nucleotide-binding loop. An intriguing protrusion (residues 367–382) is situated at one end of the platform. This protrusion contains highly conserved histidine and cysteines (H367, C373, C374 and C381), which together coordinate a Zn^{2+} ion (Fig. 1c). We denote this loop 'Zn thumb'. Its sequence is inserted between lobes 1 and 2 and is a highly conserved and characteristic feature of cGAS orthologues (Supplementary Fig. 1d), indicating an important functional role.

The cGAS–DNA–GTP–ATP complex

To reveal the structure of the activated conformation of cGAS, we co-crystallized cGAS$^{Mab21(td)}$ with a self-complementary 14-mer oligonucleotide, ATP, GTP and $MgCl_2$. To trap an activated conformation of cGAS^{Mab21} with DNA and bound nucleotides we mutated the NTase catalytic residues E200 and D202 to Q and N, respectively, thereby preventing catalysis during crystallization. The resulting transferase-deficient (td) variant is denoted cGAS$^{Mab21(td)}$. The structure of the cGAS$^{Mab21(td)}$–DNA–GTP–ATP complex was determined by molecular replacement using the coordinates for apo cGAS^{Mab21} as search model. $2F_o - F_c$ and $F_o - F_c$ maps revealed interpretable density for 13 out of 14 base pairs of the dsDNA duplex and for both nucleotides bound at the active site (Supplementary Fig. 2). The structure was refined at 3.1 Å resolution, resulting in a model with good *R*-factors and stereochemistry (Supplementary Table 1).

DNA is bound along the platform between the spine on one side and the Zn thumb on the other side (Fig. 2a). cGAS binds DNA predominantly by sequence-independent interactions to both phosphate-sugar backbone strands along the minor groove (Fig. 2b, c). Hereby, cGAS binds seven nucleotides at the core of the platform, which are recognized by at least eleven residues via specific side- and/or main-chain contacts. In addition to the phosphate and sugar contacts, two arginine fingers (R150 and R192) are inserted into the minor groove, additionally stabilizing the interaction in a fairly sequence-independent manner. Besides binding to the array of conserved positively charged residues at the bottom of the platform, DNA is also bound by the spine and the Zn thumb. The continuous helix of the spine in apo-cGAS^{Mab21} is interrupted in the DNA complex and a DNA backbone phosphate is bound at the central kink of the spine helix. On the other side of the platform, the Zn thumb contacts the DNA backbone near the major groove. We do not see close, direct polar contacts between Zn thumb and DNA, but do not want to rule out water-mediated interactions here (Supplementary Fig. 2a).

The Zn thumb does not substantially change conformation or location between apo and DNA-bound cGAS. It seems to be a rather rigid element, in which the zinc ion serves as a structural stabilizer of the protruding loop, similar to Zn^{2+} in regulatory domains of RIG-I-like receptors[25]. The location of the Zn thumb at the backbone near the major groove suggests that it may assist in binding to B-form DNA. In support of this, we do not see a substantial perturbation of the bound DNA from canonical B-form DNA.

Altogether, our structure suggests a specific recognition of B-form dsDNA by cGAS through an extended B-DNA binding platform and

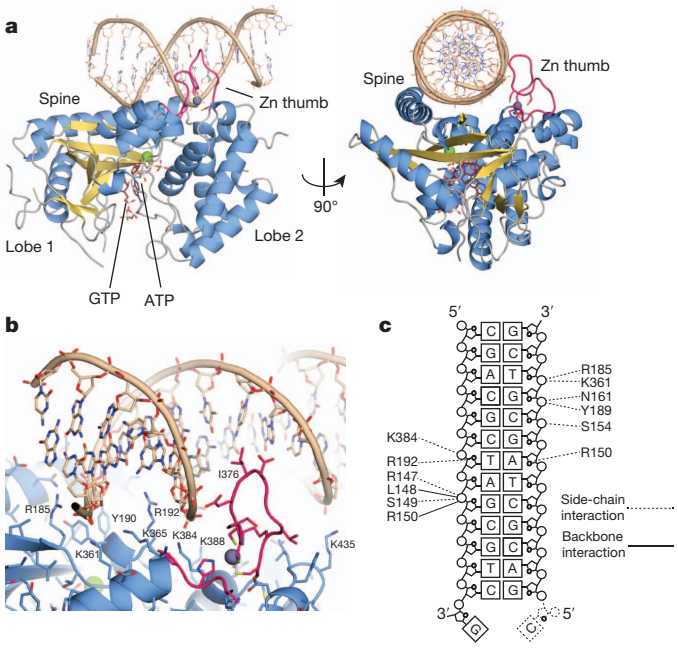


Figure 2 | The cGAS^{Mab21}–DNA–GTP–ATP complex. **a**, Side and top views of cGAS^{Mab21} (colour code of Fig. 1b) in complex with dsDNA (brown), GTP and ATP (ruby stick models). DNA binds along the platform between spine and Zn thumb. **b**, Close-up view of the DNA binding site with selected annotated residues. DNA is bound mainly via the minor groove. A notable exception is the Zn thumb near the major groove. **c**, Schematic representation of DNA–cGAS contacts.

flanking 'Zn thumb' across both lobes of the enzyme. The observed mode of binding is consistent with the key role of cGAS in sensing very different types of DNA in a sequence-independent manner[18,23].

Structure–function analysis

To validate the structural results, we mutated several conserved positively charged residues at the DNA-binding platform of human cGAS, two active site residues, two zinc ligands in the Zn thumb, or the entire Zn thumb and tested for nucleotidyl-transferase activity *in vitro* by thin-layer chromatography (TLC) (Fig. 3a). cGAS produces a product that migrates approximately in the range of c-di-AMP synthesized by DisA[26], consistent with formation of a dinucleotide. The conserved active site residues of NTases (human E225+D227; porcine E200+D202 and human G212+S213) are essential for *in vitro* activity of cGAS^{Mab21}. Moreover, mutation of conserved positively charged residues at the centre and flanking regions of the platform (K173+R176 and K407+K411) either diminish or abolish activity, in accordance with this site being important for DNA sensing. Finally, disruption of the zinc-binding site of the thumb (human C396+C397, Zn thumbless) abolishes DNA-induced NTase activity *in vitro*, highlighting the functional importance of the conserved Zn thumb in DNA binding.

To test the effect of active site, platform and thumb mutations in living cells, we measured the transactivation of an IFN-β promoter reporter by transiently expressing human cGAS variants in HEK293T cells that stably expressed murine STING (Fig. 3b). Induction of IFN-β by cGAS^{Mab21} (human cGAS$^{155–552}$) in these cells is only moderately reduced compared to wild-type cGAS, showing that the Mab21 domain structurally addressed in this study is the catalytic active functional core of the sensor. The activity of full-length cGAS was abolished when residues of the NTase active site were mutated (E225Q/A+D227N/A or G212A+S213A). Mutating charged platform residues (K173A+R176A; K407A+K411A) substantially reduced the activity of cGAS in living cells. Likewise, disrupting the zinc-binding site of the thumb (C396A+C397A, Zn thumbless) severely compromised cGAS activity. These data validate the *in vitro* biochemical data and emphasize the importance of the structure-derived motifs and elements in living cells.

To see whether Zn thumb and conserved platform surface residues are important for dsDNA binding and activity, we performed electrophoretic mobility shift assays (Fig. 3c). Both porcine and human wild-type cGAS^{Mab21} bind efficiently to dsDNA and, surprisingly, also to dsRNA (Supplementary Fig. 3a, c). The mutations in platform and thumb either did not affect DNA/RNA binding under these conditions, or reduced but did not abolish it (Supplementary Fig. 3b). However, both mutants fail to show DNA-stimulated activity under conditions where they still bind DNA, and dsRNA fails to stimulate activity under conditions where it binds robustly to the protein (Supplementary Fig. 3c, d). Thus, although these analyses validate the functional relevance of the DNA binding platform and Zn thumb on activating cGAS, they suggest that DNA or RNA interactions per se are not sufficient to activate the enzyme, indicating for instance the necessity for a precise DNA-induced structural switch.

NTase and DNA induced structural switch

To reveal the mechanism of activation of cGAS by DNA, we first analysed the NTase mechanism. We see clear electron density for two nucleotide triphosphate moieties (Supplementary Fig. 2b). The two bases partially stack in an approximately 90° rotated orientation and inserted into a hydrophobic/aromatic pocket, sandwiched between

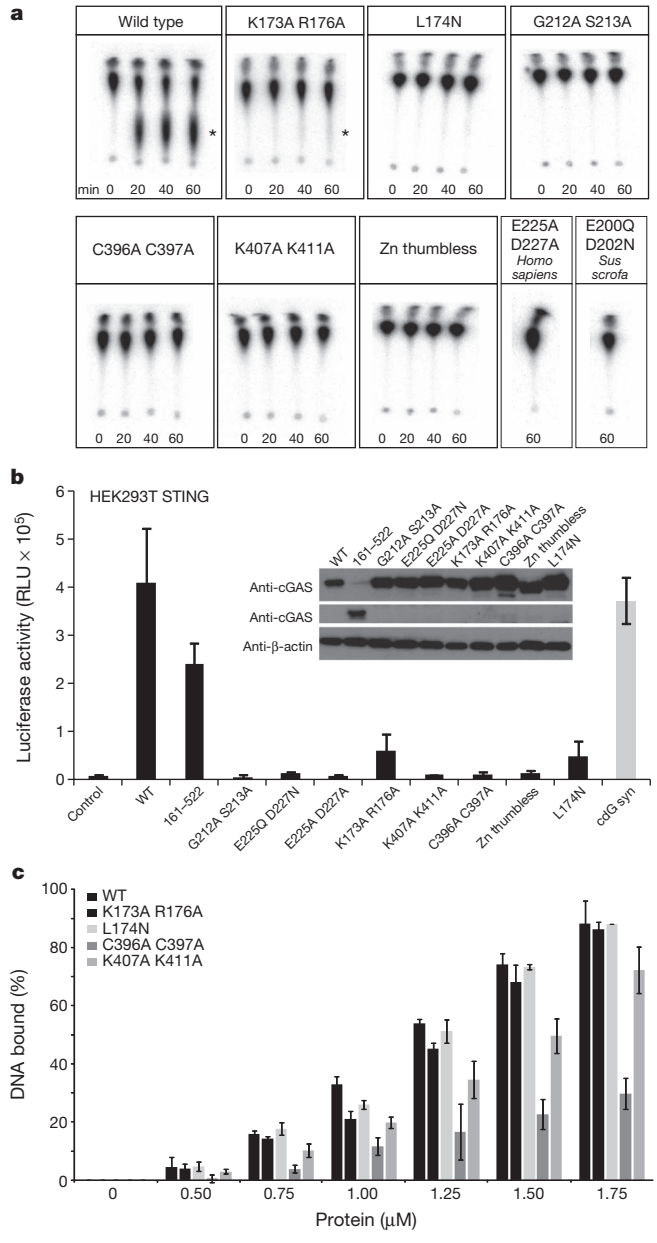


Figure 3 | Platform and Zn thumb are involved in dsDNA-dependent activity. a, NTase assays performed with different cGAS^{Mab21} mutants (2 μM) in presence of 3 μM dsDNA (50-mer). Human wild-type cGAS^{Mab21} (positive control) synthesizes dinucleotide, DNA binding site mutant K173A+R176A show reduced activity. K407A+K411A DNA binding site mutant, C396A+C397A Zn thumb mutant, Zn thumbless, L174N structural switch mutant, active site mutants E200Q+D202N of porcine cGAS^{Mab21} and E225A+D227A and G212A+S213A of human cGAS^{Mab21} are inactive. The asterisk indicates the dinucleotide product. **b**, IFN-β stimulation of cGAS mutants in HEK293T cells stably expressing murine STING. HEK293T cells were transfected with plasmids encoding indicated constructs along with the IFN-β promoter reporter plasmid pIFN-β-GLUC. Luciferase activity is plotted: mean ± s.d. ($n = 3$). Both full-length and the crystallized region (cGAS^{Mab21} human 155–522) induce IFN-β promoter transactivation. Active site mutations (G212A+S213A and E225Q/A+D227N/A) abolish IFN-β stimulation. DNA-binding site mutants (K173A+R176A, K407A+K411A), Zn thumb mutants (C396A+C397A, Zn thumbless) and structural switch mutant (L174N) either reduce or abolish IFN-β stimulation. Empty vector was used as negative control whereas cyclic-di-GMP synthase (cdG syn) expressing vector was used as positive control. Inset: western blot showing wild-type and mutant protein levels with β-actin as loading control. **c**, Electrophoretic mobility shift analysis of 50-mer dsDNA (0.2 μM) bound to cGAS^{Mab21} mutants at indicated concentrations. Plotted bars, mean ± s.d. ($n = 3$). Whereas K407A+K411A DNA binding site mutant and C396A+C397A Zn thumb mutant show slightly reduced but not impaired affinity to dsDNA, no detectable binding change was observed with the other mutants.

I298 (lobe 1) and Y413 (lobe 2). The current resolution of the diffraction data does not allow us to unambiguously determine which base is adenine and which guanine. Binding of R353 at nucleobase 1 (the 'receiving substrate' of NTases) near O6 and N7 would argue for this being guanine. In general, nucleobase 1 (interpreted as guanine here) is in hydrogen bonding distance to S355, S357 and T186, suggesting that this nucleotide is specifically recognized. In contrast, we do not observe direct hydrogen-bonding contacts of the protein to nucleobase 2 (the 'transferred' nucleotide in NTases; interpreted as adenine here). Nevertheless, this recognition might be mediated via water molecules such as in 3′ terminal uridylyl transferases[27].

The structure provides a mechanism for attack of nucleotide 1 on nucleotide 2, consistent with the mechanism of other NTases, for example, CCA adding enzyme[28]. The triphosphate chain of nucleotide 2 is well coordinated via S188 (lobe 1), S412 (lobe 2) and Mg^{2+} bound to E200 (Q in $cGAS^{Mab21(td)}$) and D202 (N in cGAS $^{Mab21(td)}$). As a consequence, the relative orientation of lobes 1 and 2 is important for the phosphate coordination of nucleotide 2. In our conformation, the α-phosphate of nucleotide 2 is well placed and oriented to promote nucleophilic attack of the sugar 2′ OH from nucleotide 1 to form the 2′-5′ linkage (Fig. 4a, see ref. 29). The attacking OH of nucleotide 1 is polarized and activated by D296, consistent with the conserved features of NTases[24]. A second Mg^{2+} could be important for this catalytic step. However, distinct localization will require higher resolution.

cGAS is proposed to form a cyclic-dinucleotide, which would require a second catalysis step and an additional attack of the OH of nucleotide 2 on the phosphate of nucleotide 1. Such an attack will require an almost 180° flip of the sugar moiety of nucleotide 2 to place its α-phosphate appropriately. In principle this is possible: in the course of our studies we determined the crystal structure of $cGAS^{Mab21}$ bound to UTP in the absence of DNA and do observe an appropriate flip of the sugar moiety (Supplementary Fig. 4). In any case, our structure satisfactorily explains the catalysis of formation of a specific (at present linear) dinucleotide by cGAS, but formation of a cyclic dinucleotide needs to be addressed in future studies.

To reveal a potential activation mechanism of cGAS, we superimposed apo-cGAS, $cGAS^{Mab21}$–UTP and $cGAS^{Mab21(td)}$–DNA–GTP–ATP complex (Fig. 4b, c and Supplementary Fig. 5a, b). We used $cGAS^{Mab21}$–UTP because UTP binding orders the β-sheets on lobe 1 and we can also visualize conformational changes specifically induced by dsDNA rather than the nucleotides.

Although UTP binding to cGAS ordered to some extend the nucleotide-binding loop in the active site, it did not substantially change the overall structure and active site geometry of cGAS (Supplementary Fig. 5b). In contrast, DNA phosphate binding to the spine (Fig. 4b) triggers a substantial structural switch in the spine helix (Fig. 4c) that closes lobes 1 and 2 and rearranges the active site loop, allowing magnesium coordinating of E200 to position and activate nucleotide 2.

To test the role of this DNA-induced structural switch we mutated human L174 to N. L174 (porcine L148) is repositioned in response to DNA binding to stabilize the nucleotide-binding loop, but does not directly bind DNA or NTPs (Supplementary Fig. 5c). Although L174N shows fairly normal DNA binding (Fig. 3c and Supplementary Fig. 3b), it lacks DNA-stimulated cGAMP synthetase activity *in vitro* (Fig. 3a) and shows decreased interferon stimulation in cells (Fig. 3b). Thus, the structural and biochemical data suggest that cGAS is activated by a DNA-induced structural switch that rearranges the NTase active site.

Conclusion

Here we provide the structure and mechanism of activation of the cytosolic DNA sensor cyclic-GMP-AMP synthase that readily explain the synthesis of a linear dinucleotide intermediate by cGAS in response to DNA binding. The backbone binding of a canonical B-DNA by cGAS is consistent with a broad specificity innate immune PRR for cytosolic DNA and the structural elements of cGAS such as the positioning of residues involved in minor-groove binding, arginine fingers and the Zn thumb suggest that cGAS specifically responds to B-form DNA. This might explain the function of other innate immune DNA sensors to detect non-canonical DNA structures, such as DAI[5]. A structural switch transmitted by proper B-form DNA binding to the active site could also explain the lack of activation by dsRNA or in mutants that still bind DNA: slightly different conformations of RNA-bound or DNA-bound mutant cGAS would not trigger robust cGAMP synthesis as even small differences in the active site geometry can strongly affect catalytic rates of enzymes.

In future, it will be important to address the specificity for other DNA structures in the activation of cGAS in more detail to see which types of DNA structures can activate cGAS. It will also be important to investigate additional requirements for efficient DNA sensing *in vivo*, because although shorter dsDNA molecules can stimulate $cGAS^{Mab21}$ *in vitro*, DNA larger than 50-mer is required for efficient IFN stimulation *in vivo*[14,19]. One possibility is that fraying of shorter DNA molecules prevents efficient stimulation or that the positively charged N terminus contributes to sensing of longer DNA molecules. In addition, STING might have a direct role in DNA binding in a larger context *in vivo*[19], although we do not see strong DNA binding *in vitro* and IFN stimulation in response to DNA in HEK293T cells in the absence of cGAS (Supplementary Fig. 6).

Interestingly, cGAS has remarkable fold similarity to the antiviral protein oligoadenylate synthase 1 (OAS1)[30,31] (Fig. 5). OAS1 synthesizes 2′-5′ linked oligoadenylate chains in response to binding to

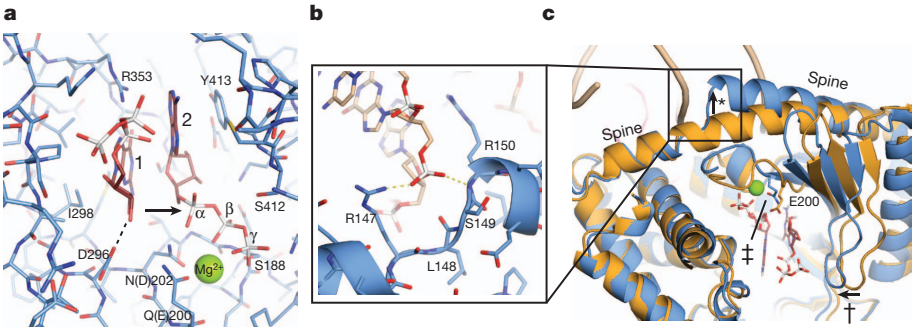


Figure 4 | NTase and DNA-induced structural switch. **a**, Close-up view of the NTase active site. Selected residues that are implicated in binding and catalysis are annotated. Both base moieties partially stack to each other and are further bound by stacking to Y413 and recognition by R353. E200 (mutated to Q in our structure) and D202 (mutated to N in our structure) bind an active site magnesium that coordinates phosphates of nucleotide 2. The attacking OH of nucleotide 1 is activated by D296 for nucleophilic attack on the α-phosphate of nucleotide 2 (arrow). **b**, Close-up view of DNA backbone phosphate binding at the spine. **c**, This DNA phosphate binding triggers a change in the spine helix (*), which allows a closure of the active site cleft (†) and repositioning of the substrate binding loop for Mg^{2+} coordination of E200 (‡).

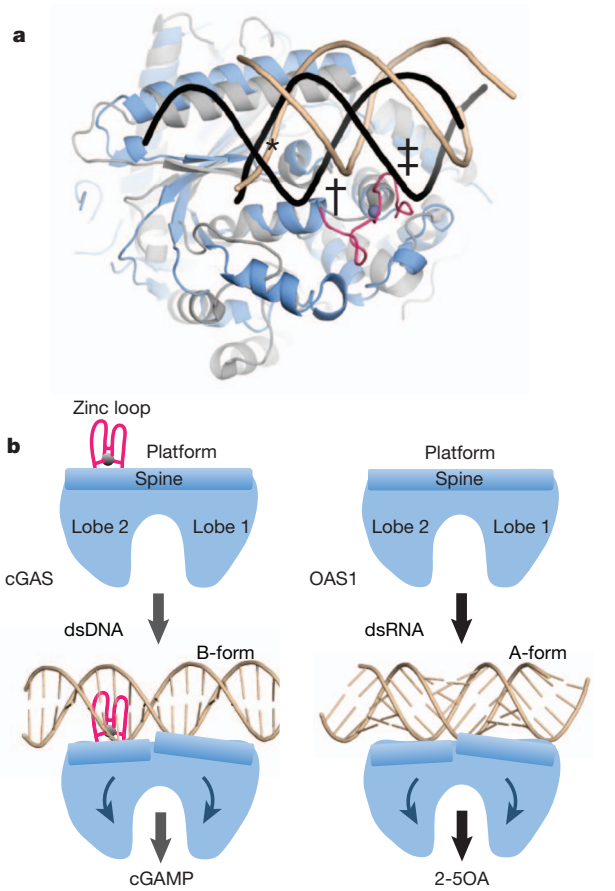


Figure 5 | Model for DNA sensing by cGAS. a, Superposition of cGAS–DNA (blue) with OAS1–RNA (grey) shows key elements for nucleic duplex selectivity. Both enzymes bind one DNA (brown)/RNA (black) backbone at the same protein site (*). The Zn thumb specifically recognizes the position of the second DNA strand in B-form (†).However, it would clash with A-form RNA/DNA (‡). **b**, Unified activation model for cytosolic double-stranded nucleic acid sensing by cGAS and OAS1 NTases by a ligand induced structural switch. 2-5OA, 2′-5′ linked oligoadenylate chains.

cytosolic dsRNA. The structural similarity not only embraces the overall fold, several active site features and arrangement of lobes 1 and 2, but also certain structural elements of the platform, including the long 'spine' helix. Like cGAS, OAS1 binds dsRNA along the 'platform' and triggers a structural change that is transmitted to the active site[31]. However, whereas OAS1 is activated by A-form RNA, cGAS is activated by B-form DNA. The Zn thumb in cGAS, missing in OAS1, probably acts as a molecular 'ruler' to specifically trigger activation in response to B-form but not A-form nucleic acids (Fig. 5). Despite these differences, cGAS shows a structural switch induced by dsDNA that is very similar to that of OAS1 induced by dsRNA[31] (Fig. 5). Thus, our results structurally unify dsDNA and dsRNA sensing by cGAS and OAS1 NTases, respectively, in the innate immune system and suggest that both processes are evolutionarily connected. *Note added in proof*: After submission of the revised version of this manuscript, Gao *et al.*[32] reported related structures of cGAS and its complexes with DNA and nucleotides.

METHODS SUMMARY

Proteins were produced in *Escherichia coli* and purified by affinity, ion exchange and size exclusion chromatography. Apo, UTP- and DNA–ATP–GTP-bound $cGAS^{Mab21}$ and its catalytic inactive form were crystallized by hanging or sitting drop vapour diffusion. The structure of apo $cGAS^{Mab21}$ was determined by single-anomalous dispersion phasing on selenomethionine derivatized protein. The other structures were determined by molecular replacement using apo $cGAS^{Mab21}$ as search model. NTase assays were performed by thin layer chromatography and phosphor imaging. DNA and RNA binding were assessed by electrophoretic mobility shift assays. Analysis of cGAS mutants in living cells were performed in HEK 293T cells stably expressing full-length murine STING and transfected with an IFN-β promoter reporter plasmid.

Full Methods and any associated references are available in the online version of the paper.

Received 21 March; accepted 20 May 2013.
Published online 30 May 2013.

1. Rathinam, V. A. K. & Fitzgerald, K. A. Cytosolic surveillance and antiviral immunity. *Curr. Opin. Virol.* **1,** 455–462 (2011).
2. Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140,** 805–820 (2010).
3. Keating, S. E., Baran, M. & Bowie, A. G. Cytosolic DNA sensors regulating type I interferon induction. *Trends Immunol.* **32,** 574–581 (2011).
4. Krug, A. Nucleic acid recognition receptors in autoimmunity. *Handb. Exp. Pharmacol.* **183,** 129–151 (2008).
5. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448,** 501–505 (2007).
6. Bürckstümmer, T. *et al.* An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nature Immunol.* **10,** 266–272 (2009).
7. Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J. & Alnemri, E. S. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* **458,** 509–513 (2009).
8. Hornung, V. *et al.* AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **458,** 514–518 (2009).
9. Ablasser, A. *et al.* RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nature Immunol.* **10,** 1065–1072 (2009).
10. Chiu, Y. H., Macmillan, J. B. & Chen, Z. J. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* **138,** 576–591 (2009).
11. Yang, P. *et al.* The cytosolic nucleic acid sensor LRRFIP1 mediates the production of type I interferon via a β-catenin-dependent pathway. *Nature Immunol.* **11,** 487–494 (2010).
12. Kim, T. *et al.* Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. *Proc. Natl Acad. Sci. USA* **107,** 15181–15186 (2010).
13. Zhang, Z. *et al.* The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nature Immunol.* **12,** 959–965 (2011).
14. Unterholzner, L. *et al.* IFI16 is an innate immune sensor for intracellular DNA. *Nature Immunol.* **11,** 997–1004 (2010).
15. Rathinam, V. A. & Fitzgerald, K. A. Innate immune sensing of DNA viruses. *Virology* **411,** 153–162 (2011).
16. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* **461,** 788–792 (2009).
17. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455,** 674–678 (2008).
18. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339,** 826–830 (2013).
19. Abe, T. *et al.* STING recognition of cytoplasmic DNA instigates cellular defense. *Mol. Cell* **50,** 5–15 (2013).
20. Burdette, D. L. *et al.* STING is a direct innate immune sensor of cyclic di-GMP. *Nature* **478,** 515–518 (2011).
21. McWhirter, S. M. *et al.* A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. *J. Exp. Med.* **206,** 1899–1911 (2009).
22. Woodward, J. J., Iavarone, A. T. & Portnoy, D. A. c-di-AMP secreted by intracellular *Listeria monocytogenes* activates a host type I interferon response. *Science* **328,** 1703–1705 (2010).
23. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
24. Kuchta, K., Knizewski, L., Wyrwicz, L. S., Rychlewski, L. & Ginalski, K. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res.* **37,** 7701–7714 (2009).
25. Cui, S. *et al.* The C-terminal regulatory domain is the RNA 5′-triphosphate sensor of RIG-I. *Mol. Cell* **29,** 169–179 (2008).
26. Witte, G., Hartung, S., Buttner, K. & Hopfner, K. P. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell* **30,** 167–178 (2008).
27. Stagno, J., Aphasizheva, I., Rosengarth, A., Luecke, H. & Aphasizhev, R. UTP-bound and apo structures of a minimal RNA uridylyltransferase. *J. Mol. Biol.* **366,** 882–899 (2007).
28. Xiong, Y. & Steitz, T. A. Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template. *Nature* **430,** 640–645 (2004).
29. Ablasser, A. *et al.* cGAS produces a 2′-5′-linked cyclic dinucleotide second messenger that activates STING *Nature* http://dx.doi.org/10.1038/nature12306 (30 May 2013).

30. Hartmann, R., Justesen, J., Sarkar, S. N., Sen, G. C. & Yee, V. C. Crystal structure of the 2′-specific and double-stranded RNA-activated interferon-induced antiviral protein 2′-5′-oligoadenylate synthetase. *Mol. Cell* **12,** 1173–1185 (2003).
31. Donovan, J., Dufner, M. & Korennykh, A. Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc. Natl Acad. Sci. USA* **110,** 1652–1657 (2013).
32. Gao, P. *et al.* Cyclic [G(2′,5′)pA(3′,5′)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153,** 1094–1107 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Butryn for comments on the manuscript. We thank the Max-Planck-Crystallization facility for initial crystal screening and the Swiss Light Source, European Synchrotron Radiation Facility and the German electron synchrotron Petra III for beam time and on-site assistance. This work was funded by the National Institutes of Health (U19AI083025), the European Research Council Advanced Grant 322869, and the Center for Integrated Protein Science Munich (CIPSM) to K.-P.H., by DFG grant 3717/2-1 to G.W., by GRK1721 to K.-P.H. and G.W., by DFG grant SFB670 and ERC grant 243046 to V.H.; C.C.O.M. is supported by GRK1721.

Author Contributions F.C. crystallized and determined the structure of cGAS, performed biochemical assays, interpreted data and wrote the manuscript. T.D. crystallized and refined the DNA complex. C.C.O.M., A.A., T.D. and M.M. performed biochemical assays. G.W. performed biochemical assays, interpreted data and helped with structure determination. V.H. supervised the cell-based experiments and interpreted data. K.-P.H designed the research, helped with structure determination, interpreted data and wrote the manuscript.

Author Information Coordinates and structure factors have been deposited at the Protein Data Bank (4JLX, 4JLZ and 4KB6). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.-P.H. (hopfner@genzentrum.lmu.de).

METHODS

Constructs and cloning. The sequence encoding full-length or truncated *Homo sapiens* and *Sus scrofa* cGAS were amplified from total cDNA (courtesy of S. Bauersachs) and cloned into pIRESneo3 (Clontech) or a modified pET21 (Novagen), respectively. The mutants were generated by site-directed mutagenesis using PfuUltra (Stratagene). Zn thumbless mutant was created by replacing residues 390–405 (*Homo sapiens*) by three Gly-Ser replicates.

Protein production and purification. All proteins were produced in *E. coli* Rosetta (DE3) or B834 (DE3) strains for native or selenomethionine derivative proteins, respectively. Bacteria were grown until a D_{600} of 0.6 to 0.8 was reached and expression was induced at 18 °C for 16 to 18 h with 0.1 mM IPTG. Proteins were purified by Ni-NTA agarose resin and incubated with tobacco etch virus (TEV) protease (ratio 1:50) at 4 °C overnight to remove the 6xHis-MBP-tag. The proteins were further purified by cation exchange chromatography followed by size exclusion chromatography using a Superdex 200 column (GE Healthcare), equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl and 1 mM DTT. Purified proteins were concentrated to 10 mg ml^{-1} for crystallization. Human STING 139–379 was purified as described[33]. All purified proteins were frozen in liquid N_2 and stored at −80 °C.

Crystallization of cGAS^{Mab21}. Purified porcine cGAS (10 mg ml^{-1}) was crystallized by hanging drop vapour diffusion in 20% PEG3350 and 200 mM sodium malonate. The crystals appeared after one day at 20 °C and were flash frozen after addition of glycerol to a final concentration of 15% (v/v). The selenomethionine derivatized protein was crystallized in 100 mM Bis-Tris propane pH 6.3, 18% PEG3350 and 200 mM sodium malonate and cryo protected with 20% ethane-1,2-diol before flash freezing. UTP bound crystals were obtained by adding 20 mM $MgCl_2$ and 1:10 (v/v) of 50 mM of nucleotide in 100 mM Tris pH 7.5 to the protein before crystallization.

For crystallizing the DNA–GTP–ATP–cGAS complex 20 mM $MgCl_2$, 2 mM of both nucleotides and 14 bp dsDNA (5′-CGACGCTAGCGTCG-3′) in a molar ratio of 1:1.2 protein:DNA were added to the inactive porcine cGAS$^{Mab21(td)}$ (E200Q+D202N) (10 mg ml^{-1}). Crystals were obtained by hanging drop vapour diffusion in 50 mM sodium cacodylate pH 7.0, 2.5 mM spermine, 60 mM $MgCl_2$ and 3% (v/v) PEG 400 after one day at 20 °C. The crystals were soaked in reservoir solution containing 25% (v/v) glycerol before flash freezing.

Data collection and refinement. X-ray diffraction data of cGAS and cGAS-UTP were collected at X06SA beamline (Swiss Light Source, Switzerland) and diffraction data of the cGAS$^{Mab21(td)}$–GTP–ATP–DNA complex were collected at PetraIII beamline P14 (EMBL/DESY, Hamburg, Germany) at 100 K. The selenomethionine derivative data were collected at the selenium peak wavelength ($\lambda = 0.97961$ Å). Data processing was carried out with XDS[34]. AutoSHARP was used to locate Se sites (SAD data set) and to produce an initial solvent flattened map[35]. An initial model was built using iterative cycles of Buccaneer[36] and ARP/wARP classic[37]. The model was optimized by alternating manual building with Coot[38] and refinement using Phenix[39] against a 2.0 Å native data set. The structure of UTP-bound cGAS and the DNA–GTP–ATP–cGAS complex structure were determined using molecular replacement with Phaser[40] and optimized by manual building with Coot and refinement with Phenix or Autobuster[41]. Data collection and refinement statistics are listed in Supplementary Table 1.

NTase assays. NTase assays were performed as described in ref. 26. Reaction mixtures with the indicated concentrations of protein and DNA (40-mer: 5′-GGATACGTAACAACGCTTATGCATCGCCGCCGCTACATCC-3′, 50-mer: 5′-GGATACGTAACAACGCTTATGCATCGCCGCCGCTACATCCCTGAGC TGAC-3′) (unless indicated 50-mer dsDNA is used) or RNA (sequence as 50-mer DNA) in 0.1 M NaCl, 40 mM Tris pH 7.5 and 10 mM $MgCl_2$ were started by addition of 100 μM ATP and 100 μM GTP containing 1:600 [α^{32}P]ATP and/or [α^{32}P]GTP (3,000 Ci mmol^{-1}, Hartmann Analytic). Analysis of the reaction products was done using thin layer chromatography (PEI-Cellulose F plates, Merck) with 1 M $(NH_4)_2SO_4$/1.5 M KH_2PO_4 pH 3.8 as running buffer for the TLC plates. Assays were performed at 35 °C. The dried TLC plates were analysed by phosphor imaging (GE Healthcare).

Electrophoretic mobility shift assays. 0.2 μM of dsDNA or dsRNA (same sequences used for NTase assays) was incubated with indicated amount of purified protein for 30 min on ice. As reaction buffer 20 mM Tris pH 8.0 and 200 mM NaCl was used. Samples were separated by 1% agarose gel prepared with Gel-Red (Biotium) as suggested by the manufacturer. The gel images were analysed using ImageJ.

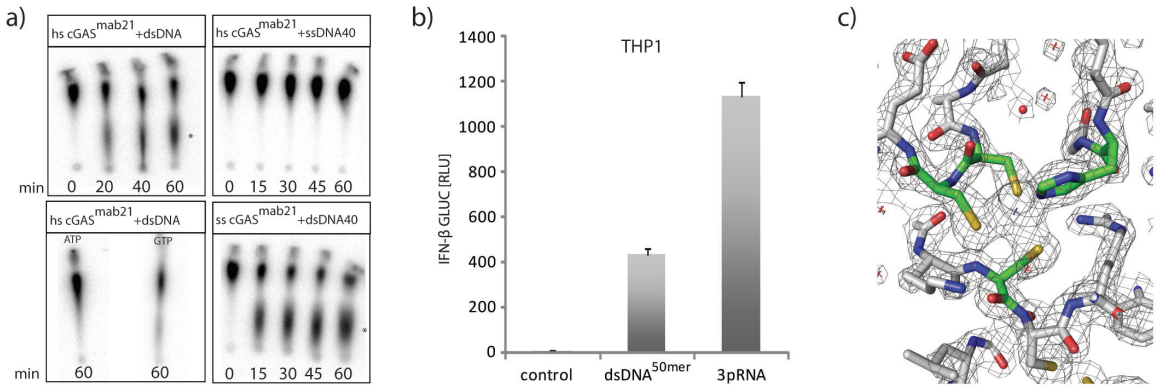
Reporter assays. HEK 293T cells stably expressing full-length murine STING (2×10^4 cells in each well of a 96-well plate) were transiently transfected with 25 ng IFN-β promoter reporter plasmid (pIFN-β-GLUC) in conjunction with 200 ng cGAS expression vectors using GeneJuice (Novagen) as indicated by the manufacturer. A codon-optimized version the diguanylate cyclase domain (83–248) of TM1788 (*Thermotoga maritima* MSB8) harbouring a point mutation (R158A) to enhance c-di-GMP production was cloned into pEFBOS to contain a carboxy-terminal haemagglutinin (HA) tag[42]. This construct (c-di-GMP-synthase) was used to induce c-di-GMP production within 293T cells upon transient overexpression, which served as positive control. 14 h post transfection luciferase activity was assessed.

THP-1 cells were stimulated with 200 ng of either 50-mer dsDNA (as in NTase assays) or tri-phosphate-RNA complexed with Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Supernatants were collected 18 h after stimulation and assayed for IP-10 production via ELISA. 90-mer DNA used is as described in ref. 19. CMA was purchased from Sigma Aldrich.

Immunoblotting. Cells were lysed in 1× Laemmli buffer and denatured at 95 °C for 5 min. Probes were separated by 10% SDS–PAGE and transferred onto nitrocellulose membranes. Blots were incubated with anti-cGAS (Sigma Aldrich), anti-phospho-IRF3 (Cell Signaling Technology) or anti-β-actin-IgG–horseradish peroxidase (HRP). Goat anti-rabbit-IgG–HRP was purchased from Santa Cruz Biotechnology.

33. Cavlar, T., Deimling, T., Ablasser, A., Hopfner, K. P. & Hornung, V. Species-specific detection of the antiviral small-molecule compound CMA by STING. *EMBO J.* **32,** 1440–1450 (2013).
34. Kabsch, W. XDS. *Acta Crystallogr.* **66,** 125–132 (2010).
35. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364,** 215–230 (2007).
36. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr.* **62,** 1002–1011 (2006).
37. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr.* **58,** 968–975 (2002).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **60,** 2126–2132 (2004).
39. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **66,** 213–221 (2010).
40. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
41. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr.* **60,** 2210–2221 (2004).
42. Rao, F. *et al.* Enzymatic synthesis of c-di-GMP using a thermophilic diguanylate cyclase. *Anal. Biochem.* **389,** 138–142 (2009).

doi:10.1038/nature12305



d)

```
ss                                  αA                                                                  β1
ss  135 GAWKLQTVLEKVRLSRHEISEAAEVVNWVVEHLLRRLQGGESEFKGVALLRTGSYYEHVKISAPNEFDVM
hs  161 GASKLRAVLEKLKLSRDDISTAAGMVKGVVDHLLLRLK.CDSAFRGVGLLNTGSYYEHVKISAPNEFDVM
mm  148 ..DKLKKVLDKLRLKRKDISEAAETVNKVVERLLRRMQKRESEFKGVEQLNTGSYYEHVKISAPNEFDVM
dr  222 LGDVLKATLDKLTIKKSERSKASRCVNEITEKVIAHLKQDTTWCADIERLRTGSYYENLKICEPDEFDVM
gg   55 GGLRLREVLSQLSLGRQDVSEASGLVNHVVSHLIQAVRGRDGSFSSIRRLGAGSYYEHVKISEPNEFDIM

ss                  β2          β3                                  αB
ss  205 FKLEVPRIQLEEYCNSGAHYFVKFKRNPGGNPLEQFLEK.EILSASKMLSKFRKIIKEEIKNIEDT..GV
hs  230 FKLEVPRIQLEEYSNTRAYYFVKFKRNPKENPLSQFLEG.EILSASKMLSKFRKIIKEEINDIKDT..DV
mm  216 FKLEVPRIELQEYYETCAFYLVKFKRIPRGNPLSHFLEG.EVLSATKMLSKFRKIIKEEVKEIKDI..DV
dr  292 LTVPVERVDIQNFDEAGAFYSIALKRHPNKHPLDKFLNEDKTIQASEMLSEFRDGVKKAVEKATDLPYKI
gg  125 LVMPVTRLQLDESDDTGAYYYLTFKRNPKEKYLNRFLDEDGKLSAFKMLEDLRRIIKEELKHIKNV..EV

ss          β4      β5          β6                                  αC      β7
ss  272 TVERKRRGSPAVTLLISKP.KEISVDIILALESK.SSWPASTQKGLPISQWLGAKVKNNLKRQPFYLVPK
hs  297 IMKRKRGGSPAVTLLIS...EKISVDITLALESK.SSWPASTQEGLRIQNWLSAKVRKQLRLKPFYLVPK
mm  283 SVEKEKPGSPAVTLLIRNP.EEISVDIILALESK.GSWPISTKEGLPIQGWLGTKVRTNLRREPFYLVPK
dr  362 KIQRKKPKCPAVTLEVTEGRKNISVDFVLGLKVHRASWPELTKDGFRVEHWLGRKVKSTMKRQPFYLVPK
gg  193 TVKRKKRGSPAITLLIKKPPAEISVDIILALEVQ.QSWPPSTQDGLNIECWLGRKVRREFRYKPIYLVAK

ss                  β8      αD          Zinc Thumb              αE
ss  340 HAKEGSGFQE.....ETWRLSFSHIEKDILKNHGQSKTCCEIDGVKCCRKECLKLMKYLLEQLKKKFGNR
hs  363 HAKEGNGFQE.....ETWRLSFSHIEKEILNNHGKSKTCCENKEEKCCRKDCLKLMKYLLEQLKERFKDK
mm  351 NAKDGNSFQG.....ETWRLSFSHTEKYILNNHGIEKTCCESSGAKCCRKECLKLMKYLLEQLKKEFQ..
dr  432 YEGIGNAEHDGVVARDSWRISFSHIEKEILKSHGHTKTCCEGREQKCCRKECLKLLKYLLQQLKNDESKS
gg  262 QNKKEKLPRG.....NTWRLSFSHIEKAMLNNHGSIKTCCESDGVKCCRKDCLKLLKYLLERLKMKHAK.

ss                  αF                  αG
ss  405 RELAKFCSYHVKTAFFHVCTQDPHDNQWHLKNLECCFDNCVAYFLQCLKTEQLANYFIPGVNLFSRDLID
hs  428 KHLDKFSSYHVKTAFFHVCTQNPQDSQWDRKDLGLCFDNCVTYFLQCLRTEKLENYFIPEFNLFSSNLID
mm  414 .ELDAFCSYHVKTAIFHMWTQDPQDSQWDPRNLSSCFDKLLAFFLECLRTEKLDHYFIPKFNLFSQELID
dr  502 NKMSSFCSYHAKTTLLQACASRGTDIEWAYSELANCFQQLLEDFVKHLKNHQLPNFFIPSHNLLH..HVS
gg  326 .ELEKFCSYHVKTAFFHSCVTWPSDTDWQHENLEHCFQKYLEYFLRCLQDSQLPHFFIPQYNLLS..LDD

ss          αH
ss  475 KPSKEFLSKQIEYERNNGFPVFW..
hs  498 KRSKEFLTKQIEYERNNEFPVFDEF
mm  483 RKSKEFLSKKIEYERNNGFPIFDKL
dr  570 TSNCDFLAKEIEFQLNNTFPIFS..
gg  393 KASNDFLLGQINYQLNNGFPVFHQV
```

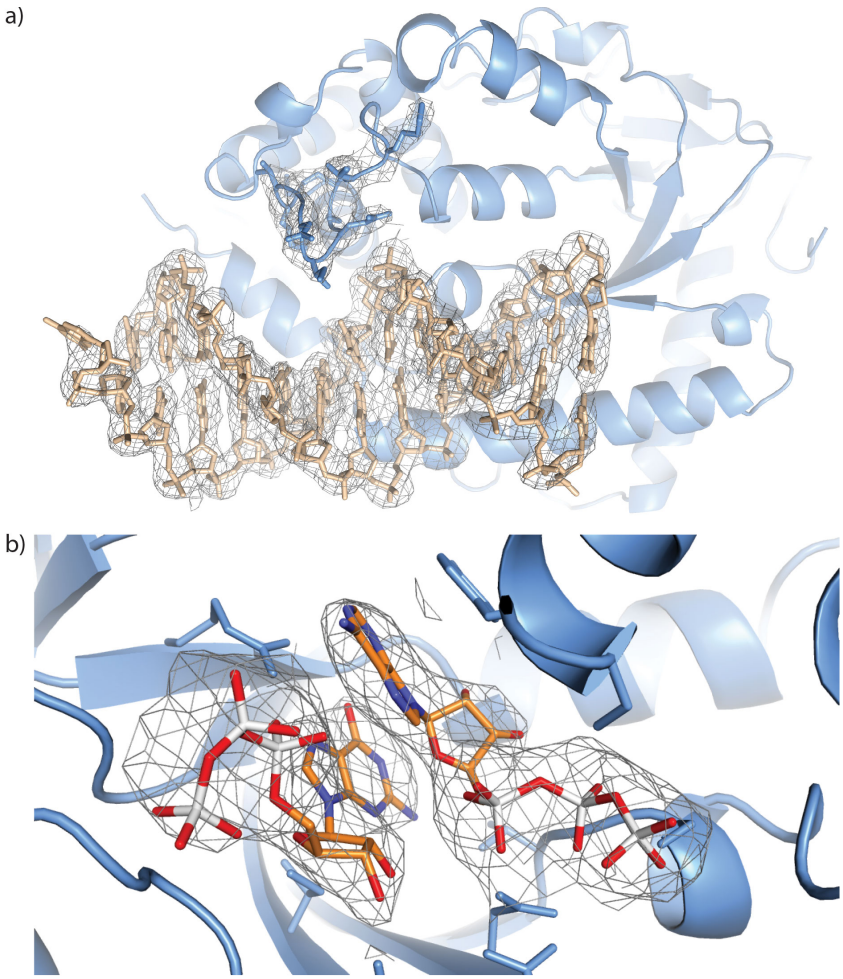
Supplemental Figure S1: Activity assays, electron density and cGAS sequence alignment

a) Activity assays with human and porcine cGAS^{mab21} in the presence of ATP, GTP and ATPα^{32}P. Upper panel left: 2µM human cGAS^{Mab21} + dsDNA (50mer), right: 1µM human cGAS^{Mab21} + 0.5µM ssDNA (40mer). Lower panel left: 2µM human cGAS^{Mab21} + 3µM dsDNA (50mer) using only ATP/ATPα^{32}P and only GTP/GTPα^{32}P, respectively; right: 1µM porcine cGAS ^{Mab21} + 0.5µM dsDNA (40mer). The reactions were stopped at indicated time points by plotting on TLC plates. Both human and porcine cGAS^{Mab21} are activated by dsDNA while ssDNA and single nucleotides fail to induce activity.

b) IFN-β stimulation in THP1 cells by dsDNA^{50mer} or with 5'triphosphate dsRNA (3pRNA). 200ng of indicated ligand was transfected to THP1 cells along with IFN-β promoter reporter plasmid pIFN-β-GLUC. Luciferase activity is plotted: mean ± sd (n=3). The dsDNA^{50mer} used in *in vitro* assays induces interferon production in THP1 cells. 3pRNA, which induces the RIG-I pathway, is used as positive control. The negative control is without ligand.

c) $2F_o$-F_c electron density overlaid with the final model around the thumb (carbons green). The $2F_o$-F_c map is contoured at 2σ.

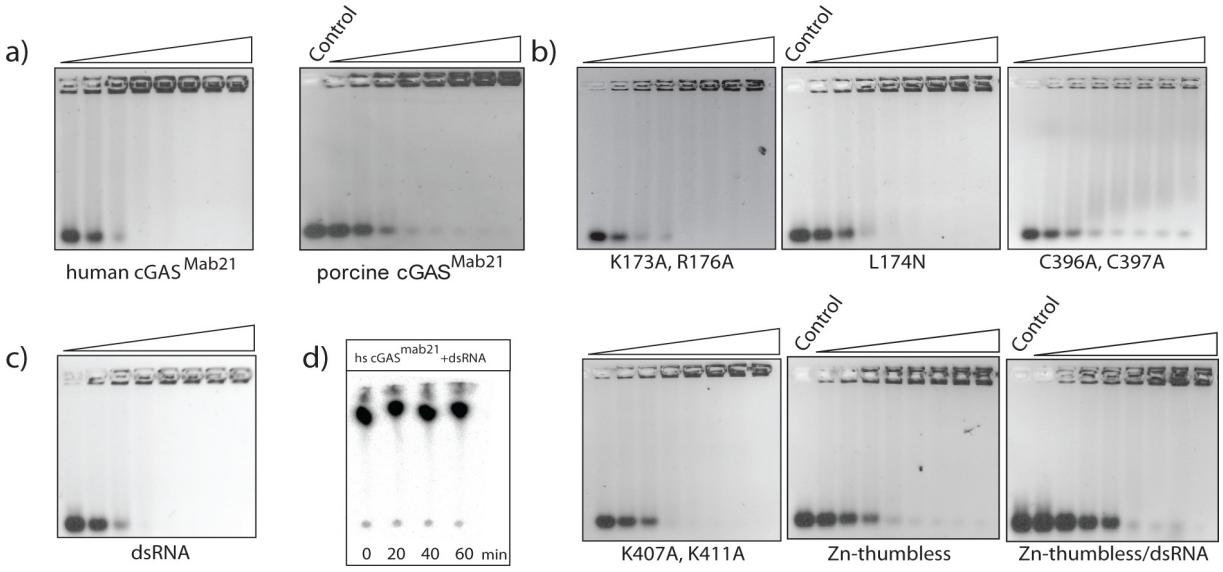
d) Structure based alignment of selected cGAS^{Mab21} sequences (abbreviations: *Sus scrofa*: **ss**, *Homo sapiens*: **hs**, *Mus musculus*: **mm**, *Danio rerio*: **dr**, *Gallus gallus*: **gg**) with highlighted conserved residues and annotated motifs. The secondary structural elements are shown on top of the alignment for the porcine homolog, color coding is analogous to Fig. 1. The squares mark residue contacts: green => DNA, magenta => GTP/ATP, cyan => Mg^{2+}, orange => Zn^{2+}. Stars denote residues that are mutated in this study.



Supplemental Figure S2: Electron densities

a) Ribbon model of cGAS$^{Mab21(td)}$ (blue) with bound dsDNA (beige stick representation). $2F_o$-F_c electron density for the DNA and the Zn-thumb residues are shown at a contour level of 1σ.

b) $2F_o$-F_c electron density around the two nucleotides in the active site (contour level of 1σ).



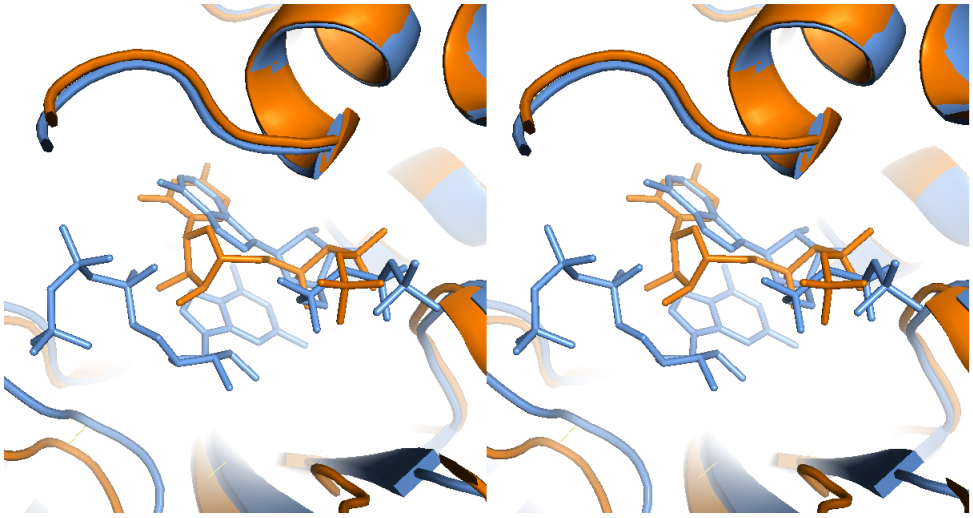
Supplementary Figure S3: Electrophoretic mobility shift and activity assays

a) Electrophoretic mobility shift analysis of dsDNA^{50mer} (0.2 µM) with human or porcine cGAS^{Mab21} (protein concentrations are 0.5, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50 µM (triangle)). Control: without protein.

b) Like a), but with human cGAS^{Mab21} mutants.

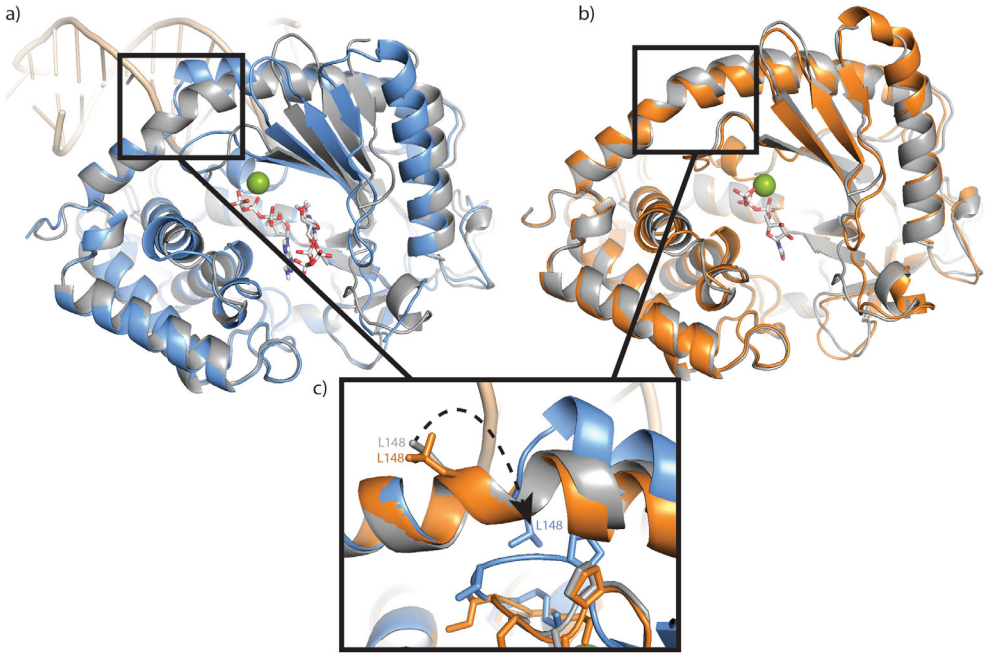
c) Like a) but binding of human cGAS^{Mab21} to dsRNA^{50mer}.

d) Activity assay with 2µM human cGAS^{Mab21} + 3µM dsRNA^{50mer}.



Supplementary Figure S4: Comparison of UTP- and ATP/GTP-bound structure of cGAS^{Mab21}

The stereo figure shows the active site of the superimposed structures of the UTP-bound state of cGAS^{Mab21} (orange) and the cGAS$^{Mab21(td)}$:DNA:GTP:ATP complex (blue), respectively. The ribose moiety is flipped in UTP as compared to ATP.

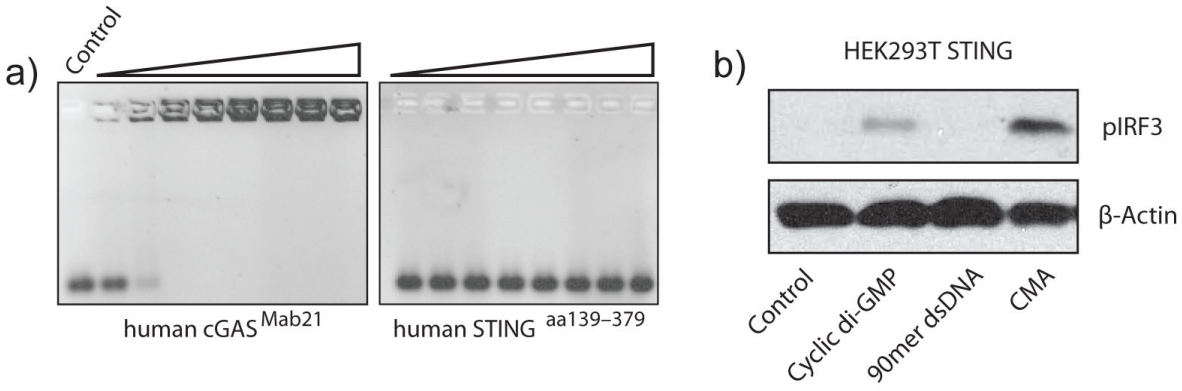


Supplementary Figure S5: Comparison of cGAS^{Mab21} structures

a) Superposition of the DNA-bound (blue) and apo (grey) cGAS^{Mab21} structures.

b) Superposition of apo (grey) cGAS^{Mab21} and UTP-bound (orange) cGAS^{Mab21}.

c) The Leu148 flip upon DNA binding. In the absence of DNA (orange UTP-bound and grey apo structures) Leu148 is solvent exposed. In the DNA binding conformation (blue), Leu148 flips and helps stabilize the nucleotide-binding loop.



Supplementary Figure S6: Comparison of STING and cGAS DNA binding and STING in vivo activity.

a) Electrophoretic mobility shift analysis of dsDNA (0.2 μM) with human cGAS^{Mab21} (left) and human STING$^{aa139-379}$ (right) (protein concentrations are 0.5, 1.0, 2.0, 3.0, 5.0, 7.0, 10, 15 μM, control: without protein). While human cGAS^{Mab21} readily binds dsDNA at a concentration as low as 0.5 μM, human STING$^{aa139-379}$ fails binding even at concentration as high as 15 μM.

b) Western blot analysis of IRF3 phosphorylation which indicates interferon stimulation. The HEK293T cells stably expressing STING were induced with cyclic-di-GMP, dsDNA^{90mer} and 10-carboxymethyl-9-acridanone (CMA). The proven STING ligands cyclic-di-GMP and CMA induce IRF3 phosphorylation while DNA^{90mer} fails, suggesting requirement of a sensor that detects DNA upstream of STING. β-actin was used as loading control.

Supplemental Table S1 Data collection and refinement statistics

	SeMet	Apo	+UTP	DNA-GTP-ATP complex
Data collection				
Space group	$C222_1$	$C222_1$	$P2_12_12_1$	$C222_1$
Cell dimensions				
a, *b*, *c* (Å)	47.5, 119.9, 140.9	47.4, 118.0, 142.6	80.6, 97.7, 107.0	86.2, 111.7, 117.6
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	47.1 (2.5)*	47.5 (2.0)*	46.9 (2.28)*	999.0 (3.1)*
R_{merge}	9.0 (44.7)*	4.7 (57.5)*	5.8 (59.1)*	9.0 (90.7)*
$I / \sigma I$	13.8 (3.3)*	20.9 (2.6)*	12.0 (2.0)*	17.0 (1.9)*
Completeness (%)	98.7 (92.0)*	99.6 (98.4)*	95.9 (93.3)*	98.6 (91.7)*
Redundancy	6.8 (6.4)*	6.4 (5.9)*	2.3 (2.3)*	6.9 (6.6)*
Wavelength (Å)	0.97961	1.00665	0.97934	0.97626
Refinement				
Resolution (Å)		45.5 - 2.0	44.6 - 2.27	68.3 - 3.08
No. reflections		51611	72224	10739
R_{work} / R_{free} (%)		18.6 / 21.3	17.0 / 20.4	25.5 / 25.8
No. atoms				
Protein		2878	5969	2899
DNA		-	-	555
Ligand/ion		54	62	73
Water		132	306	2
B-factors				
Protein		52.2	38.0	101.4
DNA		-	-	113.6
Ligand/ion		65.6	41.5	129.5
Water		45.1	38.2	95.59
R.m.s. deviations				
Bond lengths (Å)		0.008	0.009	0.004
Bond angles (°)		1.07	1.19	0.888
Ramachandran (%)				
Favored		97.4	98.6	96.0
Allowed		2.3	1.4	4.0
Outliers		0.3	0	0
PDB Accession Code		4JLX	4JLZ	4KB6

*Values in parentheses are for the highest-resolution shell.

2.2 STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF THE CELL FATE DETERMING NUCLEOTIDYLTRANSFERASE FOLD PROTEIN MAB21L1

de Oliveira Mann, C. C., Kiefersauer, R., Witte, G., and Hopfner, K.-P. (2016) Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Scientific Reports* **6**, 27498

This publication provides the first crystal structures of full-length human apoMAB21L1 and MAB21L1 in complex with CTP and thus the first structural framework for MAB21 proteins. MAB21 proteins participate in cell fate decisions and belong to the large nucleotidyltransferase fold protein family where they share considerable sequence homology with cGAS. Besides extensive characterization of MAB21 phenotypes in different model organisms and a fundamental role in development, little is known about the structure and molecular function of MAB21. This hinders amongst others the characterization of an evolutionary connection between MAB21 proteins and the cytosolic DNA sensor cGAS. This work not only revealed a high structural conservation between MAB21L1 and cGAS, but also discovered some important differences. MAB21L1 is monomeric in solution. However, in the crystal structure MAB21L1 forms a highly symmetric double-pentameric oligomer suggesting that this may be a key feature in MAB21 protein function. In the structure MAB21L1 is trapped in an inactive conformation explaining why for MAB21L1 no nucleotidyltransferase activity was detected, yet. Similar to cGAS, MAB21 proteins may require ligand interactions inducing conformational changes and proper alignment of the active site residues. Of note, co-crystallization with CTP identified a putative ligand binding site correlating with the DNA binding 'platform' in cGAS. Finally, the MAB21L1 structure provides an explanation for the implications of MAB21L2 mutations found in patients with eye malformations. Mutated residues are required for protein stabilization and mutations disrupt stabilizing interaction networks rendering a destabilized MAB21L2 protein.

Author contribution

The author of this thesis cloned, expressed and purified human MAB21L1 protein for crystallization as well as functional *in vitro* assays. She crystallized, collected the data, solved and interpreted the structures of MAB21L1 in apo as well as MAB21L1 in complex with CTP. In order to characterize the oligomeric state of MAB21L1 in solution, she performed and evaluated static light scattering experiments and in addition conducted and helped with interpreting the SAXS measurements. Further, she carried out and evaluated electro-mobility shift assay, isothermal titration calorimetry and thermal shift assay experiments. She wrote the manuscript together with G. Witte and K.-P. Hopfner.

SCIENTIFIC REPORTS

OPEN

Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1

Received: 02 March 2016
Accepted: 20 May 2016
Published: 08 June 2016

Carina C. de Oliveira Mann[1], Reiner Kiefersauer[2], Gregor Witte[1] & Karl-Peter Hopfner[1,3]

The exceptionally conserved metazoan MAB21 proteins are implicated in cell fate decisions and share considerable sequence homology with the cyclic GMP-AMP synthase. cGAS is the major innate immune sensor for cytosolic DNA and produces the second messenger 2′-5′, 3′-5′ cyclic GMP-AMP. Little is known about the structure and biochemical function of other proteins of the cGAS-MAB21 subfamily, such as MAB21L1, MAB21L2 and MAB21L3. We have determined the crystal structure of human full-length MAB21L1. Our analysis reveals high structural conservation between MAB21L1 and cGAS but also uncovers important differences. Although monomeric in solution, MAB21L1 forms a highly symmetric double-pentameric oligomer in the crystal, raising the possibility that oligomerization could be a feature of MAB21L1. In the crystal, MAB21L1 is in an inactive conformation requiring a conformational change - similar to cGAS - to develop any nucleotidyltransferase activity. Co-crystallization with NTP identified a putative ligand binding site of MAB21 proteins that corresponds to the DNA binding site of cGAS. Finally, we offer a structure-based explanation for the effects of MAB21L2 mutations in patients with eye malformations. The underlying residues participate in fold-stabilizing interaction networks and mutations destabilize the protein. In summary, we provide a first structural framework for MAB21 proteins.

The male abnormal 21 (MAB21) like proteins form a sequence related subfamily within the large and diverse family of nucleotidyltransferase (NTase) fold proteins[1]. NTases generally catalyze the transfer of a nucleoside monophosphate (NMP) from a donor nucleoside triphosphate (NTP) to an acceptor hydroxyl. A prominent member of the MAB21 like protein family is cyclic GMP-AMP synthase (cGAS), a sensor of cytosolic DNA in the innate immune system[2,3]. cGAS detects cytosolic DNA arising from intracellular bacteria, damaged mitochondria, DNA viruses and retroviruses and triggers a type I interferon response[4–6]. Upon DNA recognition, cGAS produces the second messenger, 2′-5′, 3′-5′ cyclic GMP-AMP (2′,3′-cGAMP) by two sequential NTase reactions from substrate ATP and GTP[7,8].

Besides cGAS (also known as MAB21 domain-containing protein 1 - MB21D1), MAB21 like proteins in humans include MB21D2, MAB21-like protein 1 (MAB21L1), MAB21L2 and MAB21L3[1]. Although cGAS or cGAS-like activities are present in insects, vertebrates and possibly sea anemone, the MAB21 protein was first discovered in *C. elegans* as a factor important for cell fate determination[9–11]. Mutations in *C. elegans MAB21* cause posterior-to-anterior homeotic transformation of sensory ray 6 in the male tail. The fundamental and crucial role of MAB21 during development is emphasized by additional pleiotropic phenotypes previously described and the fact that, the null allele of *MAB21* is lethal prior to hatching. MAB21 is unusually conserved throughout species and vertebrate MAB21 proteins have almost identical primary sequences.

Despite comprehensive characterization of *MAB21* phenotypes and an emerging role in cell fate decisions, the molecular function remains unknown. In vertebrates, MAB21L1 and MAB21L2 have a 94% identical amino acid sequence and exhibit similar expression patterns[12–16]. In agreement with the observations in *C. elegans*,

[1]Ludwig-Maximilians-Universität München, Gene Center and Dept. of Biochemistry, Feodor-Lynen-Str. 25, 81377 Munich, Germany. [2]Proteros Biostructures GmbH, Bunsenstraße 7a, 82152 Martinsried, Germany. [3]Center for Integrated Protein Science (CIPSM), Ludwig-Maximilians Universität München, Feodor-Lynen Str. 25, 81377 Munich, Germany. Correspondence and requests for materials should be addressed to G.W. (email: witte@genzentrum.lmu.de) or K.-P.H. (email: hopfner@genzentrum.lmu.de)

mutations of MAB21L2 result in drastic phenotypes and defects that lead to death in mid-gestation[17–19]. Studies on *MAB21L1* revealed relatively mild phenotypes in tissues where the expression of MAB21L1 differs from that of MAB21L2, namely in the lens and in reproductive organs[20]. MAB21L2 antagonizes the effects of BMP4 in *Xenopus* via interactions with the transcription factor SMAD1 and is therefore involved in regulation of the TGF-β pathway[21,22]. Moreover, MAB21L3, which only shares a sequence identity of 25%, was recently shown to be acting downstream of the Notch pathway in cell fate specifications of multiciliate cells and ionocytes[23].

Recent current whole exome sequencing projects identified four different missense mutations in the single-exon gene *MAB21L2* in eight individuals with major eye malformations[24]. In particular, an Arg51 mutation to cysteine (R51C) led to bilateral anophtalmia, intellectual disability, and rhizomelic skeletal dysplasia[25]. A mutation of the same residue to histidine (R51H) in a large multiplex family with colobomatous microphtalmia, as well as the mutation Glu49Lys was linked to coloboma. A homozygous mutation altering a different region of the protein Arg247Gln was also associated with retinal coloboma. The underlying molecular pathology of these mutations is unclear and both protein stabilizing and destabilizing effects have been proposed[24,26].

Understanding the pivotal role of MAB21 proteins during development, their pathophysiology in human disease, and their evolutionary connection to the cGAS innate immune DNA sensor is hampered because the mechanistic function, biochemical activity and interaction partners, ligands or substrates of MAB21 proteins remain elusive. Here, we provide the first crystal structure of a MAB21 protein and the structure of MAB21L1 bound to CTP. The human MAB21L1 structure exhibits a high degree of structural conservation with cGAS but features a number of differences. Our study shows that MAB21L1 would require an activating conformational change in order to possess its potential NTase activity and unveils a potential binding site for an activating ligand of this poorly understood nucleotidyltransferase. Intriguingly, the NTase contains the residues to recognize a donor NTP and active site magnesium ions, but lacks some residues that are important for NTase activity in cGAS. Finally, we provide a structure-based explanation for the effects of mutations in *MAB21L2* associated with eye malformations and show that these mutations lead to destabilization of the fold.

Results

Crystal Structure of Human MAB21L1. We crystallized native MAB21L1 in space group C2 with unit cell constants a = 167.7 Å, b = 177.6 Å, c = 115.6 Å, β = 126.5°. The native crystals contained 5 molecules in the asymmetric unit and diffracted to a limiting resolution of 3.05 Å. In order to determine experimental phases, we also crystallized selenomethionine substituted MAB21L1 (space group $P2_1$, unit cell constants a = 122.9 Å, b = 181.0 Å, c = 131.3 Å, β = 96.5° and 10 molecules per asymmetric unit). Single anomalous diffraction (SAD) data to 3.4 Å followed by phase improvement produced an interpretable electron density map that was used to build initial models for all ten copies in the asymmetric unit using non-crystallographic symmetry restrains. This initial model was used to phase the native dataset by molecular replacement. Further model building and refinement resulted in a final atomic model with good R-factors and stereochemistry (Table 1, Materials and Methods).

MAB21L1 is a two lobed, globular protein of a mixed α/β topology with approximate molecular dimensions of 64 Å × 35 Å × 48 Å. (Fig. 1a). The N-terminal lobe possesses the NTase subdomain with a five-stranded, antiparallel and highly twisted β-sheet (β1–β9) that is flanked on the surface by two long α-helices (α1/α4). The C-terminal lobe is formed by a conical four-helix bundle (α8–α13). This bi-lobal structure is stabilized by a long N-terminal 'spine' helix (α1), reaching across and spanning both lobes, as well as a central linker region composed of two helices (α5, α6). In general, the overall topology is highly similar to cGAS in its inactive form[8,27–29]. A notable difference with respect to cGAS is the lack of the zinc-thumb region which is important for DNA binding and dimerization[27,30,31]. Similarities and differences to cGAS and other members of the NTase fold protein family are described in more detail below.

Oligomerization and Interaction Interfaces. The crystal packing of MAB21L1 is unusual and consists of an intimate assembly of ten copies of MAB12L1 into a decamer, formed from two back-to-back stacked pentameric rings. The decamer has a D5 symmetry (one five-fold and five perpendicular two-fold symmetry axes) (Fig. 2a, Supplementary Fig. S1). Whereas the MAB21L1 protomers within the decamer are intimately packed, the decamers themselves pack less dense in the lattice and the interactions resemble crystal lattice interactions (Fig. 2a). The highly symmetric packing of MAB21L1 raises the possibility that oligomerization is part of MAB21L1's function – similar to e.g. cGAS - and it might be worthwhile to describe the packing within the decamer in more detail. The N-terminal lobes of each protomer are partially surface accessible and located on the outside of the pentameric rings, whereas the C-terminal lobes are to a large extent buried and mediate the contacts within the pentamer. The pentamer interface is composed of hydrophobic and some polar interactions between the N-terminal helix α1 of one monomer and the C-terminal helices α11 and α12 of the neighboring molecule (Fig. 2a). Y10 stacks with W343 and interacts with a hydrophobic patch of other nearby residues (L305, I2, A6), while N20 hydrogen-bonds with R314/E336. A citrate molecule from the crystallization condition stabilizes this interface (Fig. 2a). The citrate is coordinated by residues R344 and K340 from one monomer together with K14 and K354 from the next monomer in a very basic surface patch. In total, each pentamer buries a surface area of approximately 2500 $Å^2$.

The interface between the two back-to-back stacked pentamers is in the range of 5000–6000 $Å^2$, i.e. the decamer contains approx. 10000 Å buried surface. The pentamer stacking is mediated by contacts between opposing lobes and the linker region, both of which are highly conserved regions of MAB21L1 and L2 (Figs 1b and 2a). Salt bridges and hydrogen bonds are formed between the linker helices α6 and α5, β3 and the loop between β2–β3 as well as between the small helices α2 and α3 from the opposing monomer. Hydrophobic interactions are formed between loops connecting β2 with β3 and loop α2–α3 (Fig. 2a). The same double-pentameric packing is observed in all three of our crystals despite of changes in space groups and unit cell constants (Supplementary Fig. S1).

Data collection	Mab21L1	Mab21L1-CTP	SeMet-Mab21L1
Space group	**C2**	**$P2_1$**	**$P2_1$**
Cell dimensions			
a, *b*, *c* (Å)	167.1, 177.0, 115.1	115.0, 177.8, 134.9	122.9, 181.0, 131.2
α, β, γ (°)	90, 126.5, 90	90, 97.6, 90	90, 96.5, 90
Wavelength	1.00000	1.00000	0.97898
Resolution (Å)	50-3.05 (3.13-3.05)	50-2.55 (2.62-2.55)	50-3.4 (3.49-3.4)
R_{sym}	4.4 (115.7)	7.5 (151)	22 (112)
I/σI	22.5 (2.1)	21.7 (2.35)	8.75 (1.88)
Completeness (%)	98.8 (98.1)	99.8 (99.7)	99.2 (98.7)
Redundancy	7.0 (7.2)	13.9 (14.0)	8.9 (8.5)
Refinement			
Resolution (Å)	50-3.05	50-2.55	
No. reflections	50741	174180	
R_{work}/R_{free}	22.3/27.9	19.9/22.7	
No. of atoms (total)	13186	28115	
Protein	13117	27413	
Ligand/ion	65	440	
Water	4	262	
B-factors			
Protein	149	93	
Ligand/ion	163	109	
Water	126	77	
R.m.s. deviations			
Bond lengths (Å)	0.009	0.003	
Bond angles (°)	1.201	0.765	
Ramachandran statistics			
favoured (%)	96.3	96.8	
allowed (%)	3.3	3.0	
disallowed (%)	0.4	0.2	
PDB code	5EOG	5EOM	

Table 1. Data collection and refinement statistics. Highest resolution shell is shown in parenthesis.

In order to address the oligomeric state of MAB21L1 in solution, we performed size exclusion chromatography coupled to static light scattering (SEC-RALS) as well as small angle X-ray scattering (SAXS) experiments. Both methods show that MAB21L1 is present as a monomer in solution, at least under the assayed conditions (Fig. 2b,c). MAB21L1 elutes as a single peak with a molecular mass determined by SEC-RALS of Mw = 47 kDa, in good agreement with the theoretical mass of a monomer (Mw = 41 kDa). Molecular weight determined from the SAXS data (Mw = 41 kDa, Rg = 2.4 nm) as well as the shape of the scattering curve are only compatible with a monomeric MAB21L1 molecule in solution. In order to rule out oligomerization in the crystal due to higher salt concentrations, we performed analytical size-exclusion experiments using buffers containing different salt concentrations up to 1 M. However, MAB21L1 still remained monomeric (Supplementary Fig. S1). Although MAB21L1 is monomeric in solution, it is possible that ligand interactions may promote oligomerization, such as in the case of cGAS[30,31].

It has been proposed that MAB21L2 binds ssRNA[24]. Consistent with this, the solvent accessible electrostatic surface of the MAB21L1 monomer shows several positively-charged regions. One notable positive surface patch is a pocket at the surface adjacent to the 'spine' helix at the side of the double-pentamer (Fig. 2d). The corresponding region binds the DNA backbone in the case of cGAS. Another notable and quite extended patch is located around the five-fold symmetry axis of the pentamer (Fig. 2d). In order to test MAB21L1's affinity for oligonucleotides, we performed electrophoretic mobility shift assays (EMSAs) with dsDNA/RNA as well as ssDNA/RNA (Fig. 3a,b). MAB21L1 has a preference for ssRNA in comparison to dsRNA and ss/dsDNA in good agreement with Rainger *et al.*[24] (Fig. 3a), as can be estimated from the concentrations at half-maximal binding (ssRNA = 3.3 μM, ssDNA = 6.7 μM; dsRNA = 5 μM, dsDNA = 6 μM). The variations in the total binding of MAB21L1 to ssRNA and ssDNA are due to weak binding constants and therefore no binding curves and respective K_D values were calculated. However, the affinities for these generic DNA or RNA sequences are considerably lower in our hands than those for DNA binding to cGAS (Fig. 3b). We could also not detect any enzymatic NTase activity of MAB21L1 either in the absence or presence of these nucleic acids.

Structure of MAB21L1 Bound to CTP Reveals Possible Ligand Binding Site. In order to identify potential nucleotide substrates of MAB21L1, we performed thermal shift assays and isothermal titration calorimetry experiments using various nucleoside phosphates. We observed large increase in the melting temperature of MAB21L1 in the presence of CTP, CDP, ATP and ADP of ΔTm = 8 °C (Supplementary Fig. S2). Other

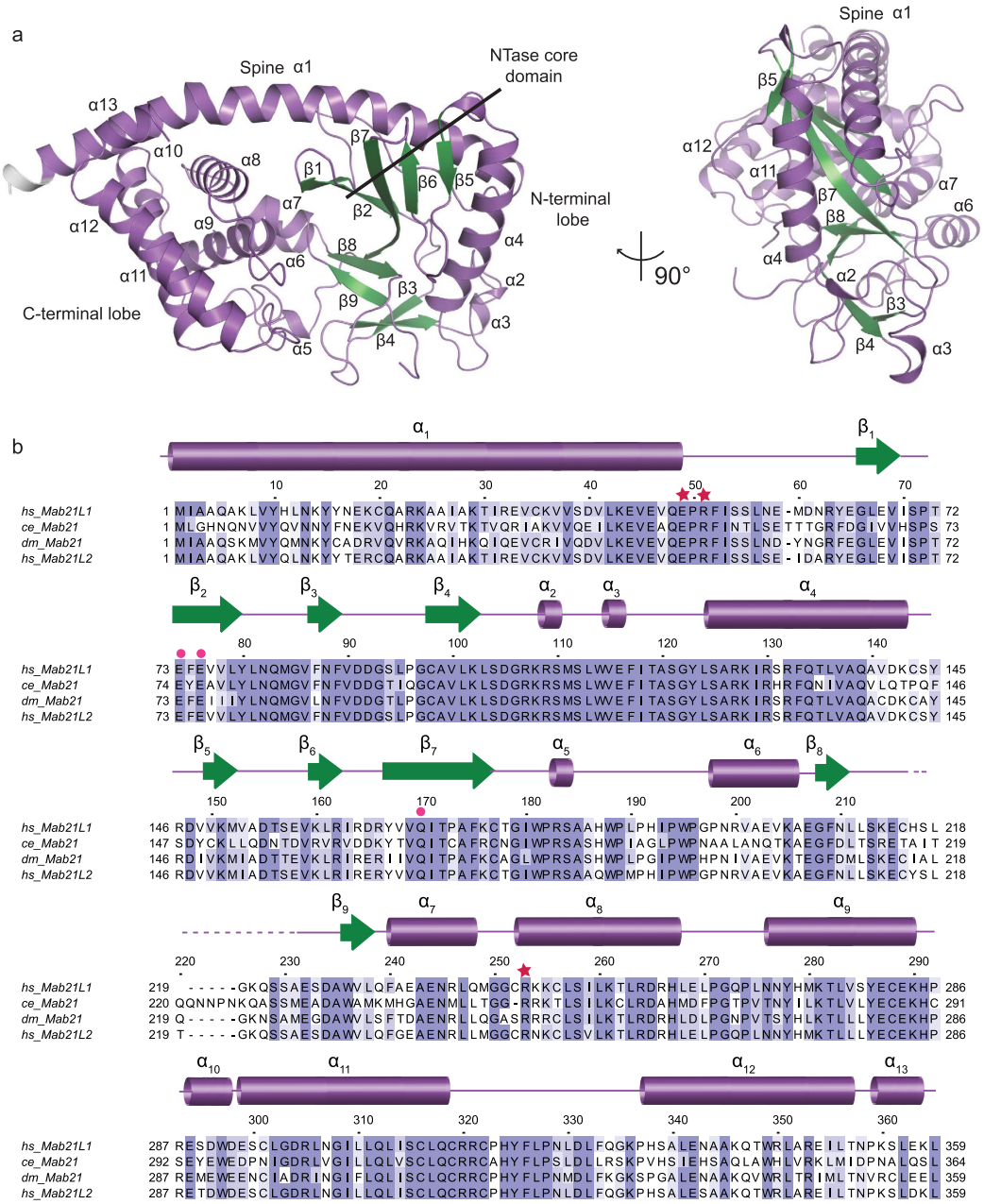


Figure 1. Structure of human MAB21L1 and Alignment with ceMAB21, dmMAB21 and hsMAB21L2. (**a**) Front and side views of the human MAB21L1 crystal structure. The model is shown as cartoon representation with annotated domains and secondary structure. (**b**) Sequence alignment of MAB21, MAB21L1 and MAB21L2 amino acid sequences (abbreviations: *Homo sapiens*: hs, *Drosophila melanogaster*: dm, *Caenorhabditis elegans*: ce). Darker shadings indicate higher conservation (BLOSUM62 conservation score). The respective secondary structure elements for the human homolog are shown on top of the alignment with dots marking the conserved active site residues usually involved in metal coordination. Red stars denote residues that have been identified in disease related mutations.

nucleotides tested, such as GTP and UTP, did not stabilize the protein and did not bind in isothermal titration calorimetry experiments (Supplementary Fig. S2). However, isothermal titration calorimetry experiments reveal strong binding of CTP, ATP and ADP with dissociation constants of $K_D(CTP) = 0.41\,\mu M$, $K_D(ATP) = 0.34\,\mu M$ and $K_D(ADP) = 3.1\,\mu M$. These data indicate a preference for a triphosphate over a diphosphate (Fig. 4c).

In order to test for Mg^{2+} -binding we performed a thermal stability shift assay in the absence and presence of magnesium. Magnesium did not stabilize MAB21L1 and this might be explained by the fact that MAB21L1 is present in an inactive conformation (Supplementary Fig. S2).

To further address the nucleotide binding, we co-crystallized MAB21L1 with both ATP and CTP in the crystallization condition and determined the resulting crystal structure of MAB21L1 to 2.55 Å resolution. In this

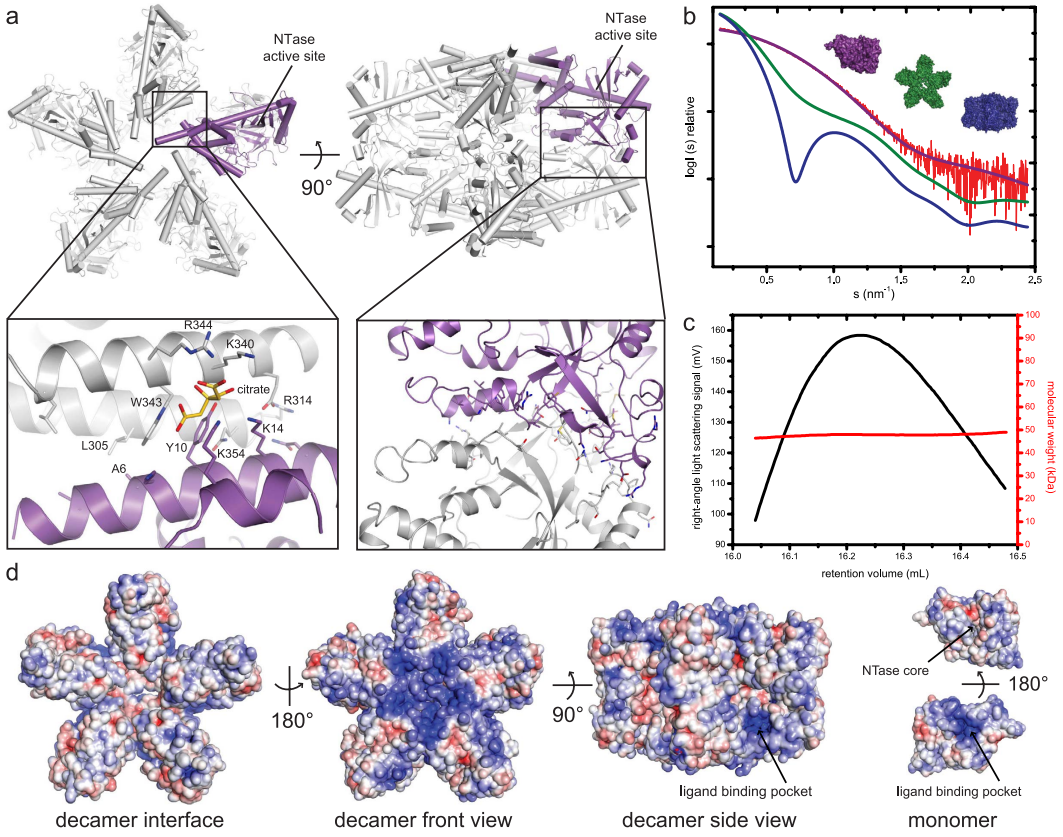


Figure 2. Characterization of MAB21L1 Oligomeric State in the Crystal and in Solution. (**a**) Front and side views of the MAB21L1 pentamer and decamer formation in the crystallographic asymmetric unit. One MAB21L1 monomer is colored in purple. The left inset shows a close-up of the hydrophobic interface between MAB21L1 molecules (purple, grey) in the pentameric ring. The right inset shows the interaction between two pentameric rings in the decameric assembly which is mainly composed of stacking interactions between highly conserved residues. (**b**) Measured SAXS data for MAB21L1 (red curve shows the measured scattering data) in comparison with theoretical scattering curves (calculated with CRYSOL) of monomeric, pentameric and decameric MAB21L1 assemblies (purple, green and blue curves, respectively) support that MAB21L1 is monomeric in solution. (**c**) Size-exclusion chromatography coupled static right-angle light scattering (SEC-RALS) of MAB21L1 shows a single peak with a constant molecular mass of Mw = 47 kDa. (**d**) Solvent accessible electrostatic surface representations of the MAB21L1 decamer interface, decamer (front view and side view), monomer (front view with NTase core and back view with ligand binding pocket) colored by charge (blue = 5 kT/e to red = −5 kT/e). The positively-charged ligand binding pocket is indicated with an arrow.

structure each protomer shows clear density for a single CTP moiety (Supplementary Fig. S3), and we did not see any density for ATP. Unexpectedly, the CTP moiety is not localized at the NTase active site, but rather bound at the positively-charged pocket on the 'platform' side of the molecule adjacent to the spine helix (Fig. 4a). The pocket coincides with a positively-charged surface patch on the 'side' of the double-pentamer (Fig. 2d). We denote this pocket 'ligand binding pocket'. CTP is well coordinated in a partially hydrophobic, partially positively-charged pocket. In the loop connecting α1-β1, R62 stacks with the cytosine base in a planar conformation (Fig. 4b). This loop is better resolved in the MAB21L1:CTP structure, compared to *apo* MAB21L1, arguing for some sort of induced fit. Further contacts of MAB21L1 with CTP include hydrophobic interactions between I31, I27, V68 and A28 and carbons of the base and sugar. Backbone interaction between L66 and Y63 and the amino group at C4 of the cytosine explain the preference of MAB21L1 for CTP and ATP, compared to UTP and GTP where no binding affinity was detected. The negative charge of the phosphate groups of the nucleotide is well and specifically coordinated by residues from α1 and α8. The α-phosphate is coordinated by K255 and R23, the β-phosphate builds salt bridges with K255 and S252 and R23, while the γ-phosphate is held in place by K24 and K248. During model building of the MAB21L1 structure with no nucleotides added, we observed an additional electron density not associated with the polypeptide chain in the ligand binding pocket (Supplementary Fig. S3). However, we were not able to identify by structural or mass spectrometric methods the nature of this low molecular weight ligand, which may be a buffer component or have co-purified with the protein from *E. coli*.

In summary, CTP binding did not obviously affect the overall conformation of MAB21L1 and the superposition of *apo* MAB21L1 (crystallized without added nucleotides) and MAB21L1 in complex with CTP reveals no major conformational changes (RMSD 0.94 Å) (Supplementary Fig. S3). Thus, it is unlikely that CTP is the physiological ligand that can induce a potential activating conformational change (see below).

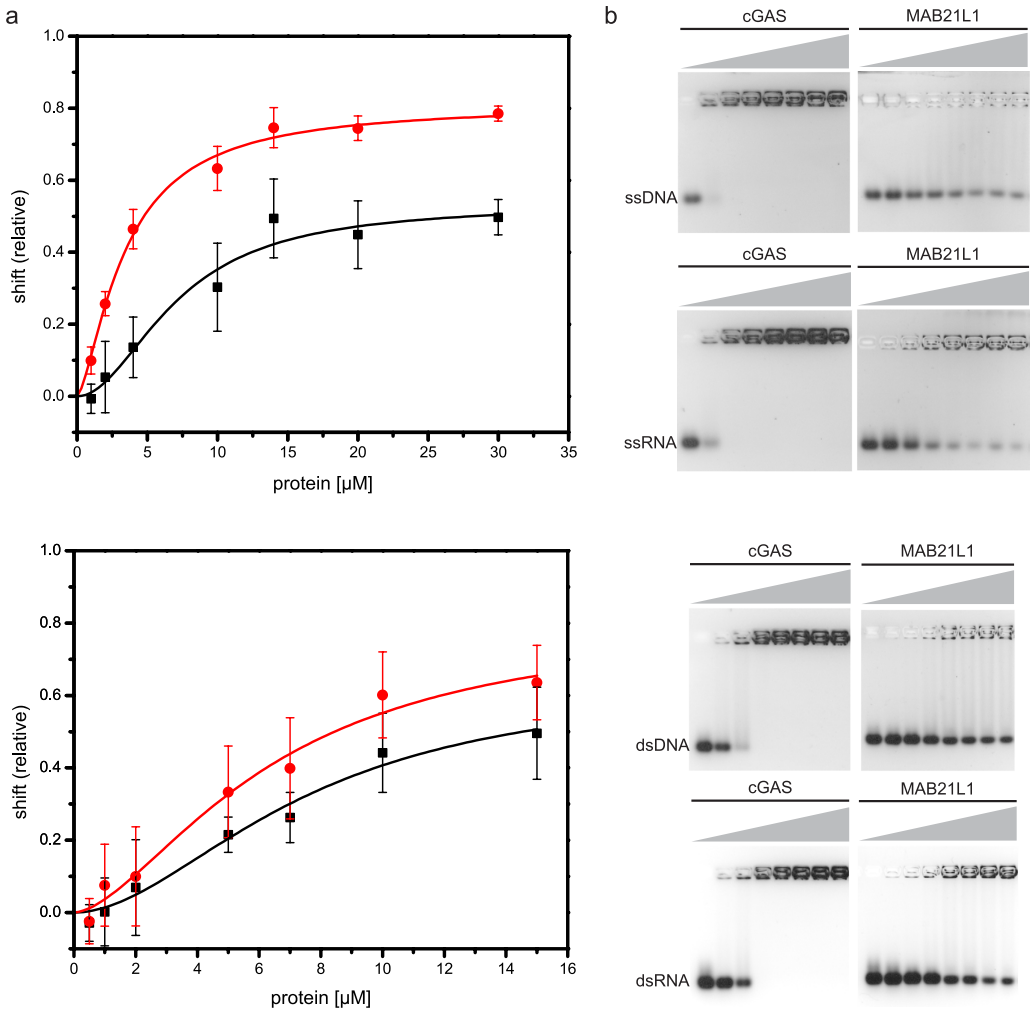


Figure 3. Characterization of MAB21L1's Affinity for Oligonucleotides. (**a**) Quantified electrophoretic mobility shift assays of 1 μM 50mer dsRNA, dsDNA and 0.5 μM, ssRNA or ssDNA with MAB21L1 (protein concentrations of 0, 0.5, 1, 2, 5, 7, 10, 15 μM for binding to dsDNA/RNA and 0, 1, 2, 10, 14, 20, 30 μM for binding to ssDNA/RNA). Intensities of unbound oligonucleotides (relative shift) were plotted as triplicates against protein concentration (upper plot: ssRNA-red, ssDNA-black; lower plot: dsRNA-red, dsDNA-black). The half-maximal binding of ssRNA is at 3.3 μM protein concentration, whereas for ssDNA it is 6.7 μM, dsRNA = 5 μM and dsDNA = 6 μM, showing that MAB21L1 has a preference for ssRNA. (**b**) Corresponding gels of electro mobility shift assays of MAB21L1 in comparison to cGAS (same conditions as before). cGAS clearly has a higher affinity for generic nucleic acids than MAB21L1.

Comparison of MAB21L1 and cGAS Active Sites. To reveal differences and similarities between MAB21L1 and cGAS in the NTase active sites, we superposed the structures of N-terminally truncated porcine *apo* cGAS (RMSD 2.2 Å, PDB code 4JLX) and porcine cGAS^{MAB21}-DNA-ATP-GTP complex (RMSD 2.8 Å, PDB code 4KB6) with the structure of the MAB21L1:CTP complex (Fig. 5). MAB21L1 shows a high degree of structural conservation to cGAS and the conserved MAB21-domain fold strongly resembles other human nucleotidyltransferases, such as OAS1, and the bacterial DncV protein (Supplementary Fig. S4). In general, many active site residues located on the α-helical C-terminal lobe – responsible for binding of the 'donor' nucleotide in cGAS and other NTases, such as OAS1 - are highly conserved (Supplementary Fig. S4). For instance, Y272 in MAB21L1 corresponds to Y413 in *ss*cGAS and to Y436 to *hs*cGAS, where it stacks with the adenine base (Fig. 5). Further conserved residues essential for correctly positioning of ATP in the active site of cGAS are *ss*cGAS E360/*hs*cGAS E383 (MAB21L1 E238), *ss*cGAS K416/*hs*cGAS K439 (MAB21L1 K275) and *ss*cGAS K391/*hs*cGAS K414 (MAB21L1 K255). Thus, at least from sequence conservation, the structure of MAB21L1 suggests that MAB21-like proteins may bind an NTP at the NTase donor position.

The N-terminal lobe binds the Mg^{2+} -triphosphate group of the donor nucleotide and specifically recognizes and orients the 'acceptor' nucleotide GTP in cGAS. Consequently, the Watson-crick and Hoogsteen edges of the guanine base are within hydrogen bonding distance of the residues *ss*cGAS T186, S355, R353 and S357 (*hs*cGAS T211, S378, R376 and S380) respectively. Whereas the donor site is partially conserved between MAB21L1 and

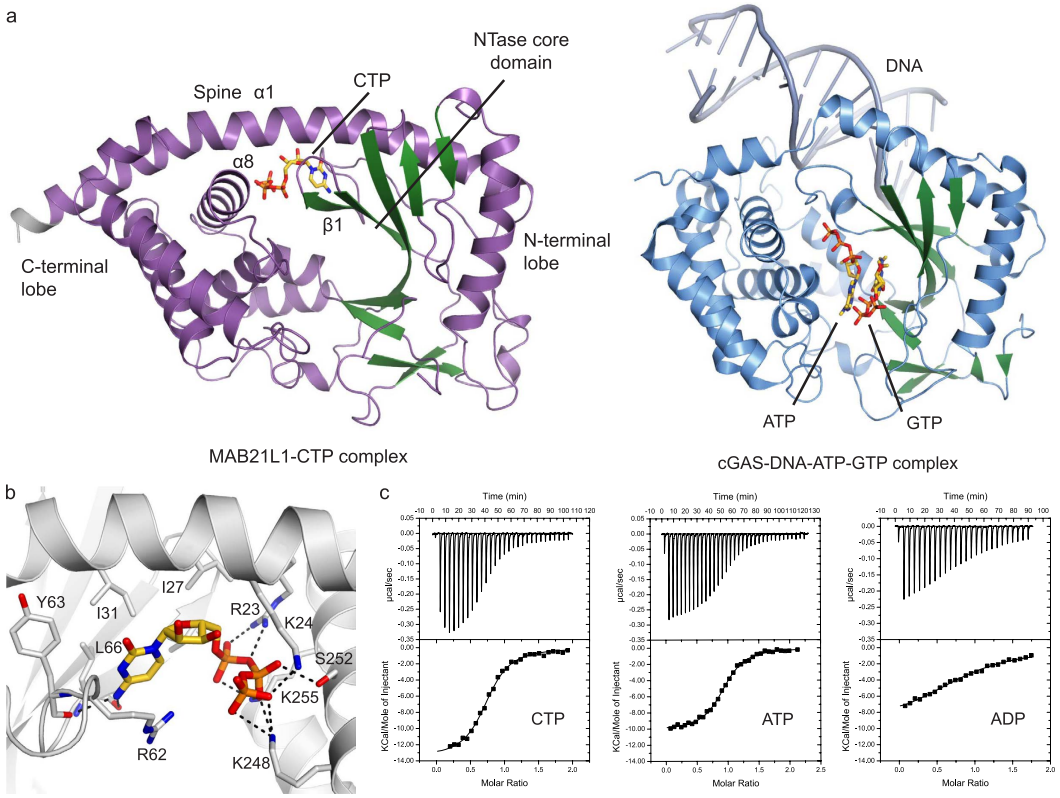


Figure 4. Structure of MAB21L1 in Complex with CTP. (**a**) Front views of the MAB21L1-CTP complex crystal structure (left) and the crystal structure of the cGAS^{MAB21}-DNA-ATP-GTP complex (right, PDB code 4KB6). (**b**) Close-up view of the CTP binding site with the bound CTP. Key residues interacting with the CTP are depicted and annotated. The CTP is bound in a positively charged pocket via interactions with the NB-loop. (**c**) Titration of MAB21L1 with different nucleotides by ITC [K_D(CTP) = 0.41 μM, K_D(ATP) = 0.34 μM and K_D(ADP) = 3.1 μM] suggest that MAB21L1 has a preference for nucleotide triphosphate coordination at this site.

cGAS, these residues, which confer cGAS specificity for the guanine base, cannot be assigned, such as *ssc*GAS T186 on the NB-loop, or are missing in MAB21L1 (A235, V231 and Q233, respectively).

A yet unanswered question is whether MAB21 proteins have any NTase activity, like e.g. cGAS or OAS. The catalytic triad on lobe I, which coordinates the two Mg^{2+} ions and polarizes the acceptor hydroxyl group in the active site of cGAS is only partially conserved in MAB21 proteins. While the presence of the Mg^{2+} ion coordinating MAB21L1 E73 and E75 (*ssc*GAS: E200 and D202/*hsc*GAS E225 and E227) suggest the existence of a functional Mg^{2+} and perhaps triphosphate coordinating site, MAB21L1 has a glutamine (Q169) at the position of *ssc*GAS D296 (*hsc*GAS D319). The carboxylate of D296 coordinates a second Mg^{2+} and helps to polarize the acceptor hydroxyl. Alanine mutations in either of the three residues or a double E200Q, D202N mutation in sscGAS lead to inactive proteins *in vitro* or in cellular assays[8]. It is unclear whether a more conserved asparagine at position hscGAS D319, similar to the Q169 of MAB21L1 still allows NTase activity.

However, in case MAB21L1 is an NTase, it would likely require a conformational change for activation. cGAS and OAS undergo conformational changes upon activation and especially part of the β1–β2 hairpin becomes a α-helical turn, which then constitutes the so called NB-loop[32]. In structures of inactive cGAS or OAS1 the loop connecting β1 and β2 is disordered in the crystal structures. Although the equivalent region of MAB21L1 is ordered in our crystal structure, it notably does not adopt the α-helical NB-loop conformation. CTP binding to MAB21L1 appears to stabilize the current, presumably inactive, conformation (Fig. 5d).

In summary, the structure of MAB21L1 exhibits similarities and differences with the structures of cGAS and OAS1, explaining why we currently fail to see NTase activity. Overall, the active site conformation observed in the crystal structure is inconsistent with catalytic activity. Residues that bind the donor NTP in the active site of NTases as well as coordinate the di-metal center are to a considerable extent conserved to cGAS, OAS proteins and other 3′-specific nucleotidyltransferases, such as poly(A) polymerase (PAP) and CCA-adding enzyme (CCA)[32].

Human Missense Mutations of MAB21L2 Mapped on MAB21L1 Structure are Responsible for Protein Stability. A whole exome sequencing project revealed four different MAB21L2 missense mutations in patients with ocular coloboma. Out of 38 candidate genes for eye malformations, MAB21L2 was the only one altered by mutations. Previous studies attempted to explain the drastic phenotype of these mutations based on alignments of MAB21L2 with cGAS[24]. However, only the mutation R247Q, located in the C-terminal lobe on α8,

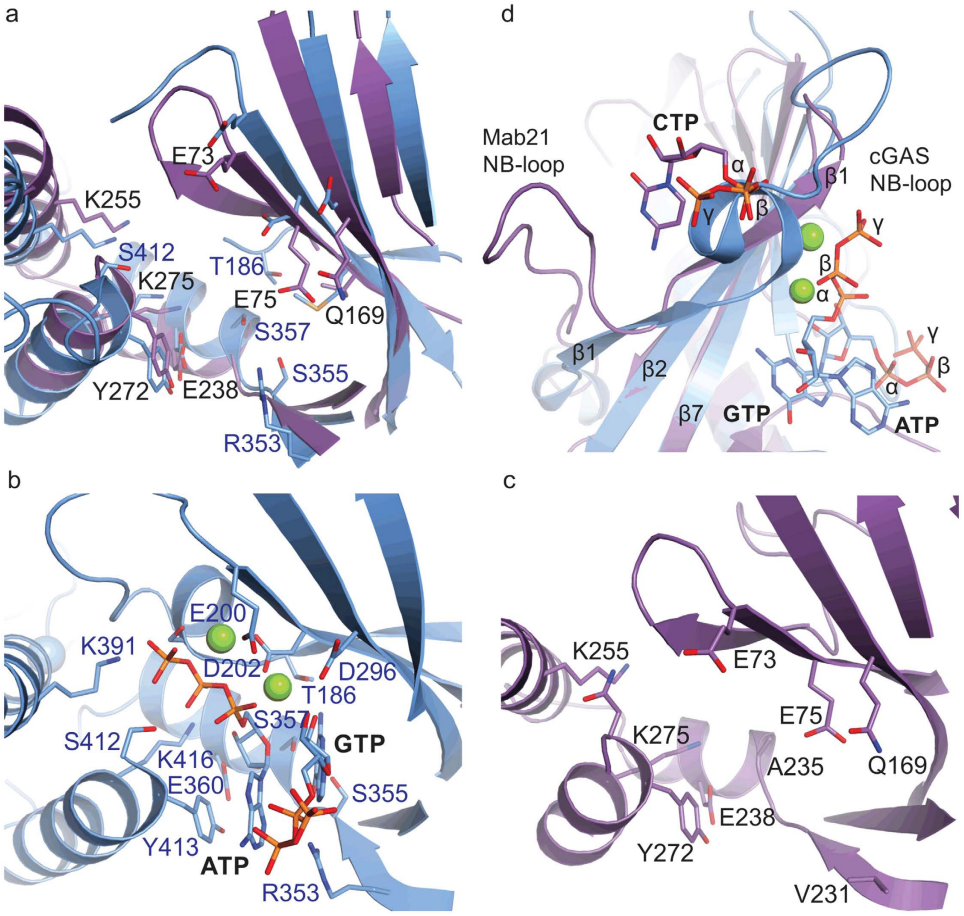


Figure 5. Superposition of MAB21L1 Active Site with Structurally Conserved cGAS. (**a**) Close-up views of the NTase active sites of human MAB21L1 (purple) superposed with porcine cGAS (blue, PDB code 4JLX). (**b**) The corresponding part of the porcine cGAS-DNA-ATP-GTP (blue, PDB code 4KB6) and (**c**) human MAB21L1 structure in the same orientation. Residues responsible for magnesium ion coordination in the cGAS-DNA-ATP-GTP structure (PDB code 4KB6) Q200 and N202 are mutated in the figure to E200 and D202, respectively, and a second magnesium ion was added to the catalytic site of cGAS for reason of clarity. Selected residues of MAB21L1, which are conserved to cGAS residues implicated in catalysis, are annotated. (**d**) Close-up view of nucleotide binding sites in the cGAS-DNA-ATP-GTP and MAB21L1-CTP complexes. CTP binding to MAB21L1 prevents the NB-loop rearrangements required for activation and therefore access to the catalytic pocket.

is structurally conserved in cGAS and even OAS1 (Supplementary Fig. S4) and previous work disagreed whether the patient derived mutations increased or decreased protein stability[24,26].

All of the described amino acids mutated in MAB21L2 are conserved in MAB21L1 (Fig. 1b), so our crystal structure offers a good model to reveal the molecular basis for this disease. We therefore mapped the mutations R51C/H, E49K and R247Q on our MAB21L1 structure, in order to shed light on the possible effects on the protein stability (Fig. 6a). In our MAB21L1 structure R247 forms a salt bridge with E288 on the α10-turn. This salt bridge is disrupted by the mutation of R247 to glutamine and consequently may lead to decreased stability of the protein. Residues E49 and R51 (Mutated to K and C/H, respectively) are located on the beginning of loop α1-β1. They are unique to MAB21 proteins and are highly conserved (Fig. 1b). Based on our MAB21L1 structure, E49 and R51 participate in a network of salt bridges with E115 (Fig. 6a). E49 and R51 stabilize the large loop between β4 and α4 by anchoring the loop via interactions to E115. This loop and E115 are a highly conserved element in MAB21L1 and L2, indicating that destabilization caused by the mutations of E49 to lysine and R51 to cysteine or histidine inactivate MAB21L2.

To experimentally address the effects of the mutations on the protein stability we performed comparative fluorescence thermal shift assays. The thermal stability of MAB21L1 was tested in comparison to MAB21L1 R247Q and MAB21L1 R51C. MAB21L1 R51C and MAB21L1 R247Q showed significantly decreased melting temperatures, which indicates a decreased stability of the mutated protein (Fig. 6b). The mutant MAB21L1 R247Q showed an even more drastic decrease in melting point stability loss as R51C. The fact that R247Q is also conserved in cGAS (*ss*cGAS R383 and *hs*cGAS R406) suggests that this residue plays a more fundamental role in protein

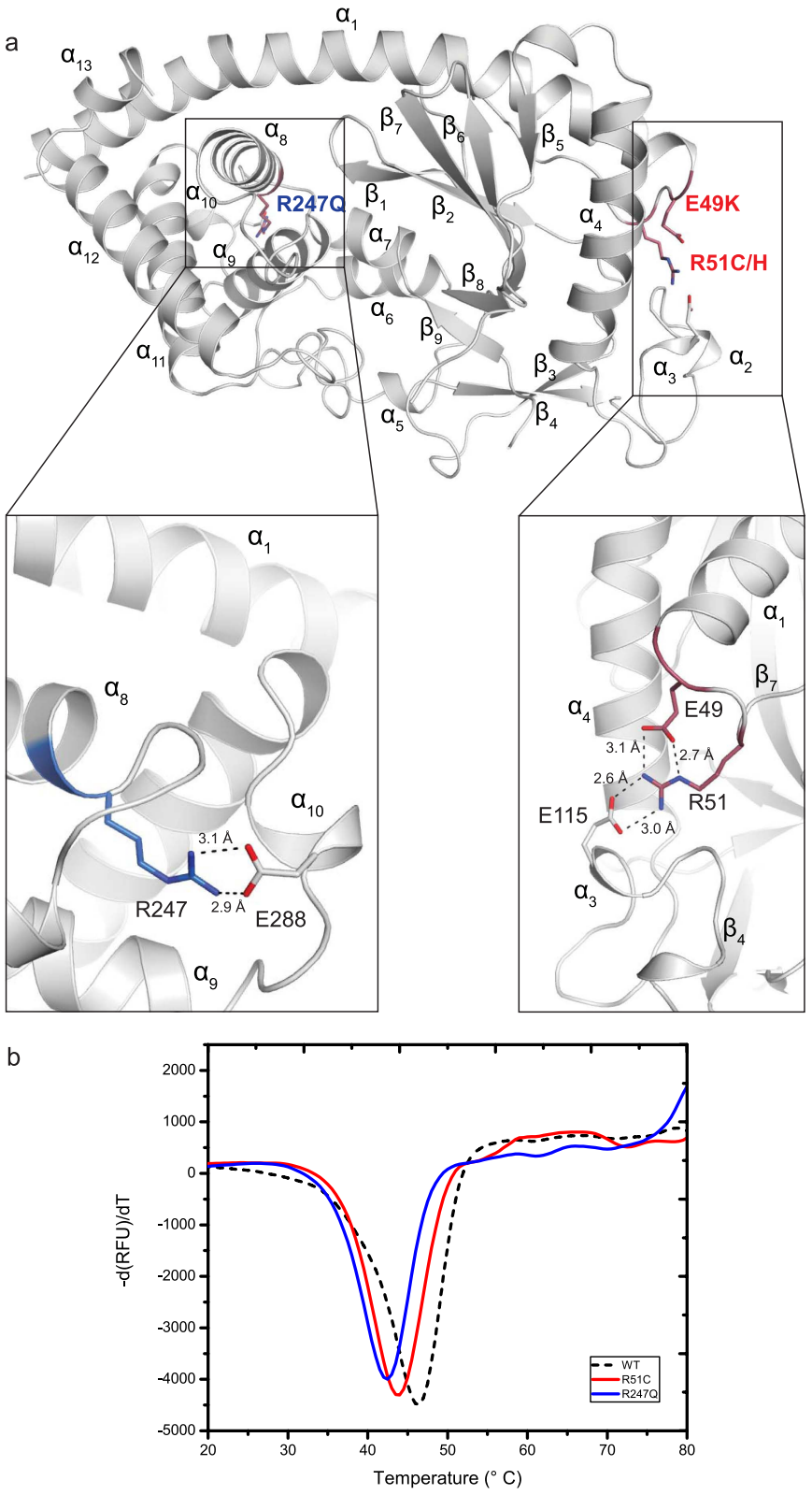


Figure 6. Patient Mutations on MAB21L2 Affect Protein Stability. (**a**) Amino acids of MAB21L2 mutated in patients with eye malformations are depicted in blue (R247Q) and red (E49K, R51C/H) in the MAB21L1 structure. The insets show close-up views of the salt bridges formed by the mutated residues. R247 forms a salt bridge with E288, while E49 and R51 build a network of salt bridges with E115. Mutation of these stabilizing interactions will affect the correct MAB21L1 fold. (**b**) Thermal shift assay derivative melt curve plots of MAB21L1 (black), MAB21L1 R51C (red) and MAB21L1 R247Q (blue). The disease related mutations in MAB21L1 all show a decreased melting temperature supporting the idea that the mutations destabilize the protein.

stability (Supplementary Fig. S4). In summary, we conclude that the MAB21L2 mutations associated with ocular coloboma lead to a protein with reduced stability because critical fold-stabilizing bonding networks are disrupted.

Discussion

MAB21 family proteins are implicated in cell fate decision processes in metazoans. Whereas MAB21 is required for pattern formation of the male tail in nematodes, MAB21 is sufficient and required for development of the dorso-ventral axis in vertebrates[10,33,34]. MAB21 proteins act downstream of the TGF-β signaling pathway by antagonizing the effect of BMP4 in *Xenopus* embryo ventralization and in *C. elegans* MAB21 is negatively regulated by cet-1 (vertebrate BMP4)[22]. Direct interactions of MAB21L2 with SMAD1, a nuclear transducer of the TGF-β pathway, suggest that MAB21L1 could be a transcriptional repressor[21]. Several studies in the model organisms *Xenopus*, zebrafish and mouse revealed that both genes have similar expression patterns in the developing eye, mid- and hindbrain, neural tube and branchial arches[13–15,35]. In particular, recent studies in zebrafish detected *MAB21L2* transcripts in the ciliary marginal zone of the retina. Similarly, expression of *MAB21L3* in *Xenopus* was specific for multiciliate cells and ionocytes associating MAB21 proteins with cell fate specification of cilia[23,26].

The biochemical activity and physiological role of MAB21 proteins remain to be established. As with many other NTases there are three ligand or substrate molecules to be considered, which make identification of the biochemical activity and biological function of these types of molecules challenging. NTases transfer a NMP from a donor NTP onto a hydroxyl group of an acceptor molecule. A third molecule required for development of catalytic activity is a potential activating ligand. In the case of cGAS, these molecules are ATP, GTP and double-stranded DNA. In the case of MAB21 proteins, neither the substrates of a NTase reaction nor a potential activating ligand have been described, if such molecules exist at all. Our structural studies show that MAB21L1 resembles cGAS in the unliganded inhibited state and show that any NTase activity would require a conformational change similar to that DNA induces in cGAS[8,27]. It is of course possible that MAB21 proteins exert their functions via a predominantly structural role, e.g. by binding other proteins or nucleic acids such as mRNAs or micro-RNAs, and do not possess any catalytic activity at all. For instance, it has previously been proposed that MAB21L2 interacts with mRNA[24]. We also see some preferential but moderate binding to generic ssRNA with MAB21L1 and the positive electrostatic surface potential would be consistent with nucleic acid binding. Generic nucleic acids bind, however, with considerably less affinity to MAB21L1 than to cGAS[27]. This reduced affinity suggests that MAB21 binds perhaps a specific sequence or structure, if any. Of note in this context, potential functional link to micro-RNAs in the case of MAB21L3 has been proposed, although there are no experimental data for a functional or physical interaction at this point[23]. Apart from nucleic acids, it is also possible that MAB21 proteins interact with other protein partners such as the SMAD proteins[21].

MAB21 proteins could also confer their function through a catalytic NTase activity. As stated above, MAB21L1 has an inactive conformation in the crystal structure, but a cGAS-like conformational change could in principle form a proper active site. If such a hypothetical conformational change takes place, our structure shows that residues typically associated with donor NTP binding are conserved, in particular at the C-terminal lobe 2. These include Y272, which stacks with the base, and E238, which binds to the sugar OH groups. Finally, donor NTP binding also requires the catalytic triad carboxylates on lobe I, which coordinate together with the triphosphate moiety the two active site magnesium ions. However, residues coordinating the acceptor nucleotide GTP in cGAS are not found in MAB21L1 and residues located on the NB-loop cannot be assigned in our conformation, due to the lack of a MAB21L1 structure in an active conformation. Furthermore, in MAB21L1 and L2, two out of the three catalytic triad residues are conserved (E73, E75), whereas the third triad residue (Q169) is altered. In the case of cGAS the aspartate equivalent to Q169 (*hs*cGAS D319) coordinates a second active site Mg^{2+} and could help to polarize the attacking acceptor OH. A glutamine at this position could still coordinate this magnesium ion, but may result in a reduced but perhaps not entirely abolished activity. As such, we do not want to rule out the possibility that MAB21 proteins still possess NTase activity.

An unexpected observation was that CTP was not bound at the NTase active site, even though it binds to MAB21L1 with 0.41 μM affinity and leads to a substantial increase in thermal stability. In previous studies by ourselves and other groups with related enzymes (cGAS, OAS, MiD51), co-crystallization with NTPs or soaking of crystals with NTPs resulted in binding of the NTPs at the NTase active site, although the NTPs were often bound in a catalytically inactive conformation and not sufficient to switch the enzyme from an inactive to an active conformation[27,32,36,37]. In our structure, CTP is bound to a site that recognizes activating ligands in the case of related NTases (cGAS, OAS1), and it is tempting to speculate that this site is a ligand-binding site of MAB21 proteins as well. Obviously CTP did not switch the protein into an active conformation, so CTP is unlikely a physiological activating ligand. It is possible however, that CTP mimics part of a physiological ligand. It should be noted that the sugar and base moieties of CTP are bound by the NB-loop in its inactive conformation and thus would argue against an activating structural switch in MAB21, while the triphosphate binds into a highly positively-charged pocket that may not be affected by a conformational switch. The fact that the structure of MAB21L1 without any nucleotides added shows density for an unidentified molecule in the ligand binding pocket may support the suggestion that this positively-charged site acts as a binding site for an acidic ligand. The explicit preference of MAB21L1 for a nucleoside triphosphate (over a diphosphate) at this site suggests that MAB21L1 could bind, for instance, the 5′ triphosphate of an RNA such as mRNA or precursor miRNA. This positively-charged pocket in MAB21L1 could act as an activation site but also as an inactivation site, such as the equivalent site on the bacterial nucleotidyltransferase DncV that binds 5MTHFGLU2[38]. In the folate-bound conformation of DncV residues Y117 and Q116 sterically hinder binding of nucleotides in the active site.

In our structure, MAB21L1 packs into a decamer with D5 symmetry. This highly symmetric packing is quite unusual for generic crystal lattice interactions. Although we cannot rule out that this packing is purely a result of the high concentrations in the crystal, it may have a biological function, even though it was not detected *in vitro* in solution. cGAS, for instance, was shown to cooperatively bind dsDNA through two distinct interaction

sites, which induced oligomerization required for activation[30,31]. The interaction interfaces between protomers in the double pentamer are rather small and distinct from those of the oligomers of related proteins, such as cGAS and MiD51[30,36]. The interface area between two pentameric rings, however, is substantial and mediated by a highly conserved region in MAB21L1 and L2 comprising amino acids G96-Q138 (Fig. 1b). The same sequence was shown to be involved in protein-protein interactions in related MAB21 domain proteins. Drp1 recruitment to MiD51 was abolished when a single amino acid was mutated in this region[36]. In addition, residues differing between the 94% identical MAB21L1 and MAB21L2 proteins do not affect the predicted oligomer interface. The extreme surface conservation of MAB21 proteins throughout species further argues for potential oligomerization and/ or interaction partner binding features. Therefore, the oligomerization of MAB21 could be of biological importance and should be considered in future studies.

Finally, our structure allows a molecular interpretation of the MAB21L2 mutations in ocular coloboma since the underlying residues and their interaction partners are conserved between L1 and L2. Interestingly, we find that the mutated residues are not located on the surface but rather stabilize the structure of MAB21L1/L2. Consistently, the mutations reduce the thermal stability of MAB21L1 *in vitro*, suggesting that the disease is caused by destabilized MAB21L2. Of note, the mutations E49K and R51C/H help fold a highly conserved area in MAB21L1/L2 that in our structure mediates the pentamer-pentamer interactions. In MiD51 the same loop is responsible for protein-protein interactions[36]. Therefore, the phenotype caused by mutations E49K and R51C/H could be due to disruption of possible protein-protein interactions with a yet unknown partner, that might be required for MAB21L1 to fulfill its function. Furthermore, these mutations were previously shown to disrupt MAB21L2 ability to bind ssRNA, consistent with the locations of the affected residues close to the RNA binding cleft in OAS1[24].

In line with the mutations observed for MAB21L2, a recent study reports severe effects of a the Cys246Leufs*18 frameshift mutation in MAB21L1[39]. This mutation leads to a potentially disrupted structural integrity of the whole alpha helical arrangement in the C-terminal lobe, also affecting the mapped intramolecular interaction R247:E288 in MAB21L2 mutation R247Q. The fact that the MAB21L1 mutation causes an additional phenotype compared to MAB21L2 argues for an important role of both proteins in the organism.

In summary, we provide a first structural framework for the highly conserved MAB21 family of nucleotidyltransferase fold proteins.

Methods

Protein Expression and Purification. The full length human MAB21L1 gene, purchased from GENEART with codon optimization for expression in *E. coli,* was cloned into a modified pET21 vector with an N-terminal His-MBP-tag. Site-directed mutagenesis was performed using the Quikchange (Stratagene) protocol with the respective primers. Expression was performed in *E. coli* BL21 Rosetta (DE3) and B834 (DE3) strains for native and selenomethionine labeled proteins, respectively. Cells were induced with 0.1 mM IPTG after reaching an OD_{600} 0.5–0.6 and proteins were expressed at 18 °C for 18 h. Harvested cells were resuspended in lysis buffer (50 mM Tris, 300 mM NaCl, 25 mM imidazole, 2 mM β-mercaptoethanol, 5% v/v glycerol) and disrupted by sonication. MAB21L1 was purified by nickel-affinity chromatography and the His-MBP-tag was subsequently removed by TEV-protease cleavage (ratio 1:30) at 4 °C over night in buffer A (30 mM Tris-HCl, 100 mM NaCl, 2 mM DTT, pH 7.0). The protein was then further purified by cation exchange chromatography step using a HiTrap SP HP column (GE Healthcare) at pH 7.0. The protein was eluted with a linear gradient up to 50% buffer B (30 mM Tris, 1 M NaCl, 2 mM DTT, pH 7.0). Fractions containing MAB21L1 were pooled and loaded on a Superdex 75 size exclusion chromatography column (GE Healthcare) using 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 mM DTT as running buffer. MAB21L1 fractions were concentrated (2–3 mg/ml) before being flash-frozen in liquid nitrogen for storage at −80 °C. MAB21L1 mutants were expressed and purified accordingly. The final crystallization buffer for selenomethionine derivatized protein contained 2 mM TCEP (Tris(2-carboxyethyl) phosphine).

Crystallization of MAB21L1. Co-crystallization of MAB21L1 with a mixture of CTP/ATP (5 mM final concentration) was performed by hanging drop vapor diffusion in 0.1 M MES pH 5.5 and 1 M tri-sodium citrate. 2 μL protein/nucleotide mix were added to 1 μL reservoir solution from a total reservoir volume of 250 μL in the well. Native crystals grew for one week at 20 °C before being mounted on a Free Mounting System (FMS, Proteros Biostructures GmbH). Liquid surrounding the crystal was removed in the humidified gas stream with a glass capillary. TMAO (trimethylamine oxide) was supplied as a cryo-protectant to the protein crystal with the PicoDropper during a humidity gradient. The crystal was then flash-frozen by the quick rotational replacement of the humidity nozzle with the cryo nozzle and finally stored in liquid nitrogen (for more details see also[40]). Selenomethionine derivatized protein crystals grew in 0.2 M Tri-potassium citrate pH 7.8 and 16% PEG 3350 and were cryo protected by soaking in reservoir solution containing 30% (v/v) glycerol prior to flash freezing.

Data Collection and Structure Determination. X-ray diffraction data were collected at beamline SLS X06SA (Swiss Light Source, Villigen, Switzerland) and PETRAIII beamline P13 (EMBL/DESY, Hamburg, Germany). Diffraction data were processed using XDS and XSCALE[41]. We determined the structure of MAB21L1 using a high-redundant single anomalous diffraction (SAD) dataset measured at the Se-peak wavelength. Three finely-sliced datasets of translations on one needle-shaped crystal were merged (10800 × 0.1° frames) to provide enough phase information. The self-rotation function calculated with MOLREP[42] of the CCP4 package[43] suggested the presence of an additional five-fold symmetry axis and in combination with the Matthews' coefficient we assumed either 10 or 15 copies. Phasing and density modification using the PHENIX suite[44] located 70 Se sites for the 10 copies of MAB21L1 in the ASU leading to an initial low resolution map at 3.4 Å. This map allowed partial automatic model building using BUCCANEER[45] followed by manual model building in COOT[46] and alternate refinement steps in PHENIX. The initial starting model could be used to phase the two better diffracting

native crystals using PHASER[47], leading to readily interpretable electron density. Extensive manual model building in COOT and further iterative refinements with PHENIX resulted in a final model with reasonable statistics. Data collections and refinement statistics are listed in Table 1. Figures were created with PyMOL[48].

Thermal Shift Assay. The thermal stability of MAB21L1 in presence of different ligands was analyzed by fluorescence thermal shift assays. 20 μM MAB21L1 was incubated with 5 mM of the respective ligand in 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$. After addition of SYPRO Orange (final concentration of 5×, Invitrogen) the fluorescence signal was detected during a temperature gradient between 5 °C to 100 °C at a rate of 0.5 K/30 s with one scan per 0.5 K in a real-time thermal cycler using the FRET mode (CFX96 touch/Biorad).

Isothermal Titration Calorimetry. Isothermal Titration Calorimetry was carried out at 23 °C using a VP-ITC Microcal calorimeter (Microcal, GE Healthcare). Proteins were dialyzed overnight against 20 mM HEPES pH 8.0, 150 mM NaCl, 1 mM TCEP and nucleotides were resuspended in dialysis buffer. 1.44 mL of protein in the cell with concentrations of 5–10 μM were titrated with nucleotides with 10× higher concentrations. 8 μL nucleotides was injected 25–30 times with 3.5 min intervals between injections. A background curve for each titration consisting of the titration of the nucleotide into buffer without protein was subtracted to account for heat dilution. The ITC data were analyzed using the Origin version 7 software package of the ITC instrument (Microcal).

Analytical Size-exclusion chromatography. Analytical size-exclusion chromatography with MAB21L1 was performed with a Superdex 200 increase 10/300 (GE Healthcare) in 20 mM HEPES pH 8.0, 2 mM DTT with varying concentrations of NaCl (150 mM, 500 mM, 750 mM and 1 M).

Size-exclusion chromatography coupled static light scattering. Size-exclusion coupled static right-angle light scattering (SEC-RALS) was performed using an AEKTAmicro system (GE Healthcare Life Sciences) equipped with a refractive index detector (VE 3580) and a TD270- RALS –device (Viscotek/Malvern Instruments) with a Superdex 200 10/300 size-exclusion column (GE Healthcare). BSA (66 kDa) was used as standard protein for system calibration. Analysis of data was performed with Software OmniSEC (Viscotek/Malvern Instruments).

Small-Angle X-ray Scattering. Small-angle X-ray scattering (SAXS) experiments were performed at PETRAIII beamline P12 (EMBL/DESY, Hamburg, Germany). Protein samples were purified by size-exclusion chromatography (SEC) and centrifuged prior to the measurements. All samples were monodisperse as judged from the SEC- chromatograms and dynamic light-scattering size distributions. The scattering of the SEC running buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 mM DTT) was used for buffer correction of the MAB21L1 samples which were measured at different concentrations (1.3–6.5 mg/mL). The samples did not show signs of radiation damage and the scattering data were processed in the ATSAS suite[49] as e.g. described in[50]. Guinier-plot (ln I(s) vs s^2) analysis did not show signs of aggregation of the sample and provided a radius of gyration for MAB21L1 of Rg = 2.4 nm. By extrapolation to zero angle scattering intensity (I_0) in the Guinier region $s*R_g < 1.3$ the molecular weight of MAB21L1 in solution was determined $Mw(I_0) = 49.8$ kDa using BSA (66 kDa) as a I_0-reference. This is in good agreement with the molecular weight determined from the Porod volume Mw(porod) = 41 kDa. Theoretical scattering curves of crystallographic coordinates were calculated using CRYSOL[51].

Electrophoretic Mobility Shift Assays. 1 μM of 50mer ds/ssDNA or ds/ssRNA (5′-GGATACGTAACAACG CTTATGCATCGCCCCGCTACATCCCTGAGCTGAC-3′), RNA same sequence as DNA) was incubated with the indicated increasing concentration of purified protein for 30 min on ice. As reaction buffer 20 mM Tris pH 8.0 and 200 mM NaCl was used. Samples were separated by a 1% agarose gel supplemented with Gel-Red (Biotium) for staining as suggested by the manufacturer instructions.

References

1. Kuchta, K., Knizewski, L., Wyrwicz, L. S., Rychlewski, L. & Ginalski, K. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic acids research* **37,** 7701–7714 (2009).
2. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
3. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339,** 826–830 (2013).
4. West, A. P. *et al.* Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520,** 553–557 (2015).
5. Gao, D. *et al.* Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341,** 903–906 (2013).
6. Li, X.-D. *et al.* Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* **341,** 1390–1394 (2013).
7. Ablasser, A. *et al.* cGAS produces a 2 [prime]-5 [prime]-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498,** 380–384 (2013).
8. Gao, P. *et al.* Cyclic [G (2′, 5′) pA (3′, 5′) p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153,** 1094–1107 (2013).
9. Kranzusch, P. J. *et al.* Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2′, 3′ cGAMP Signaling. *Molecular cell* **59,** 891–903 (2015).
10. Baird, S. E., Fitch, D., Kassem, I. & Emmons, S. W. Pattern formation in the nematode epidermis: determination of the arrangement of peripheral sense organs in the *C. elegans* male tail. *Development* **113,** 515–526 (1991).
11. Chow, K. L., Hall, D. H. & Emmons, S. W. The mab-21 gene of Caenorhabditis elegans encodes a novel protein required for choice of alternate cell fates. *Development* **121,** 3615–3626 (1995).
12. Mariani, M. *et al.* Two murine and human homologs of mab-21, a cell fate determination gene involved in Caenorhabditis elegans neural development. *Human molecular genetics* **8,** 2397–2406 (1999).
13. Wong, Y.-M. & Chow, K. L. Expression of zebrafish mab21 genes marks the differentiating eye, midbrain and neural tube. *Mechanisms of development* **113,** 149–152 (2002).

14. Mariani, M. *et al.* Mab21, the mouse homolog of a *C. elegans* cell-fate specification gene, participates in cerebellar, midbrain and eye development. *Mechanisms of development* **79,** 131–135 (1998).
15. Kudoh, T. & Dawid, I. B. Zebrafish mab21l2 is specifically expressed in the presumptive eye and tectum from early somitogenesis onwards. *Mechanisms of development* **109,** 95–98 (2001).
16. Cederlund, M. L. *et al.* mab21l2 transgenics reveal novel expression patterns of mab21l1 and mab21l2, and conserved promoter regulation without sequence conservation. *Developmental Dynamics* **240,** 745–754 (2011).
17. Lau, G. T. *et al.* Embryonic XMab21l2 expression is required for gastrulation and subsequent neural development. *Biochemical and biophysical research communications* **280,** 1378–1384 (2001).
18. Yamada, R., Mizutani-Koseki, Y., Koseki, H. & Takahashi, N. Requirement for Mab21l2 during development of murine retina and ventral body wall. *Developmental biology* **274,** 295–307 (2004).
19. Kennedy, B. N. *et al.* Zebrafish rx3 and mab21l2 are required during eye morphogenesis. *Developmental biology* **270,** 336–349 (2004).
20. Yamada, R. *et al.* Cell-autonomous involvement of Mab21l1 is essential for lens placode development. *Development* **130,** 1759–1770 (2003).
21. Baldessari, D., Badaloni, A., Longhi, R., Zappavigna, V. & Consalez, G. G. MAB21L2, a vertebrate member of the Male-abnormal 21 family, modulates BMP signaling and interacts with SMAD1. *BMC cell biology* **5,** 48 (2004).
22. Morita, K., Chow, K. L. & Ueno, N. Regulation of body length and male tail ray pattern formation of Caenorhabditis elegans by a member of TGF-beta family. *Development* **126,** 1337–1347 (1999).
23. Takahashi, C., Kusakabe, M., Suzuki, T., Miyatake, K. & Nishida, E. mab21-l3 regulates cell fate specification of multiciliate cells and ionocytes. *Nature communications* **6** (2015).
24. Rainger, J. *et al.* Monoallelic and biallelic mutations in MAB21L2 cause a spectrum of major eye malformations. *The American Journal of Human Genetics* **94,** 915–923 (2014).
25. Horn, D. *et al.* A Novel Oculo-Skeletal syndrome with intellectual disability caused by a particular MAB21L2 mutation. *European journal of medical genetics* **58,** 387–391 (2015).
26. Deml, B. *et al.* Mutations in MAB21L2 Result in Ocular Coloboma, Microcornea and Cataracts. *PLoS genetics* **11,** e1005002–e1005002 (2015).
27. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498,** 332–337 (2013).
28. Kranzusch, P. J., Lee, A. S.-Y., Berger, J. M. & Doudna, J. A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell reports* **3,** 1362–1368 (2013).
29. Kato, K. *et al.* Structural and functional analyses of DNA-sensing and immune activation by human cGAS. *PloS one* **8,** e76983 (2013).
30. Li, X. *et al.* Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39,** 1019–1031 (2013).
31. Zhang, X. *et al.* The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell reports* **6,** 421–430 (2014).
32. Lohöfener, J. *et al.* The Activation Mechanism of 2′-5′-Oligoadenylate Synthetase Gives New Insights Into OAS/cGAS Triggers of Innate Immunity. *Structure* **23,** 851–862 (2015).
33. Sridharan, J., Haremaki, T., Jin, Y., Teegala, S. & Weinstein, D. C. Xmab21l3 mediates dorsoventral patterning in Xenopus laevis. *Mechanisms of development* **129,** 136–146 (2012).
34. Wong, R. L. Y. & Chow, K. L. Depletion of Mab21l1 and Mab21l2 messages in mouse embryo arrests axial turning, and impairs notochord and neural tube differentiation. *Teratology* **65,** 70–77 (2002).
35. Cederlund, M. L. *et al.* mab21l2 transgenics reveal novel expression patterns of mab21l1 and mab21l2, and conserved promoter regulation without sequence conservation. *Developmental Dynamics* **240,** 745–754 (2011).
36. Losón, O. C. *et al.* The mitochondrial fission receptor MiD51 requires ADP as a cofactor. *Structure* **22,** 367–377 (2014).
37. Richter, V. *et al.* Structural and functional analysis of MiD51, a dynamin receptor required for mitochondrial fission. *The Journal of cell biology* **204,** 477–486 (2014).
38. Zhu, D. *et al.* Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. *Molecular cell* **55,** 931–937 (2014).
39. Bruel, A. L. *et al.* Autosomal recessive truncating MAB21L1 mutation associated with a syndromic scrotal agenesis. *Clinical Genetics* (2016).
40. Kiefersauer, R., Grandl, B., Krapp, S. & Huber, R. IR laser-induced protein crystal transformation. *Acta Crystallographica Section D: Biological Crystallography* **70,** 1224–1232 (2014).
41. Kabsch, W. Xds. *Acta Crystallographica Section D: Biological Crystallography* **66,** 125–132 (2010).
42. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *Journal of applied crystallography* **30,** 1022–1025 (1997).
43. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D:Biological Crystallography* **67,** 235–242 (2011).
44. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D:Biological Crystallography* **68,** 352–367 (2012).
45. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D:Biological Crystallography* **62,** 1002–1011 (2006).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D:Biological Crystallography* **60,** 2126–2132 (2004).
47. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40,** 658–674 (2007).
48. DeLano, W. L. The PyMOL molecular graphics system. (2002).
49. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of applied crystallography* **45,** 342–350 (2012).
50. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics* **40,** 191–285 (2007).
51. Svergun, D., Barberato, C. & Koch, M. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography* **28,** 768–773 (1995).

Acknowledgements

We thank the staffs of PETRA3 P12/P13/P14 at EMBL Hamburg/DESY and of SLS X06SA Villingen, Switzerland for excellent onsite support and beamtime allocation, Katja Lammens and Dirk Kostrewa for support with X-ray data, Claire Basquin for help with the ITC experiments, Robert Byrne for comments on the manuscript and the Hopfner group for discussions. This work was funded by grant GRK1721 (Deutsche Forschungsgemeinschaft) to G.W. and K.-P.H. and by DFG HO2489/8 to K.-P.H. C.C.O.M. is supported by GRK1721.

Author Contributions

C.C.d.O.M. conducted all experiments, and evaluated data. R.K. optimized cryo conditions on the Free Mounting Device. G.W. participated in structural analysis and SAXS evaluation. G.W. and K.-P.H. designed research. C.C.d.O.M., G.W. and K.-P.H. wrote the manuscript. All authors approved the final version of the manuscript.

Additional Information

Accession codes: MAB21L1 coordinates and structure factors have been deposited in the Protein Data Bank with the accession codes 5EOG (apo-form) and 5EOM (CTP-bound), respectively.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: R.K. declares that he is an employee of Proteros Biostructures GmbH. R.K. has invented and further developed the FMS and the PicoDropper.

How to cite this article: de Oliveira Mann, C. C. *et al.* Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci. Rep.* **6**, 27498; doi: 10.1038/srep27498 (2016).

Supplementary data

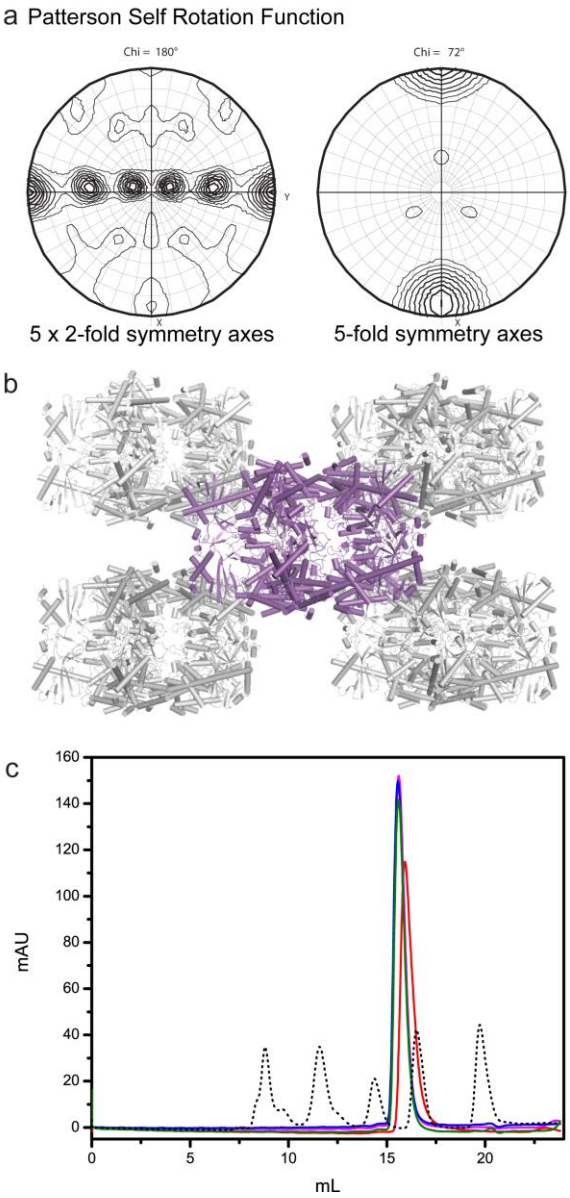
Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1

Carina C. de Oliveira Mann[1], Reiner Kiefersauer[2], Gregor Witte[1*], and Karl-Peter Hopfner[1,3*]

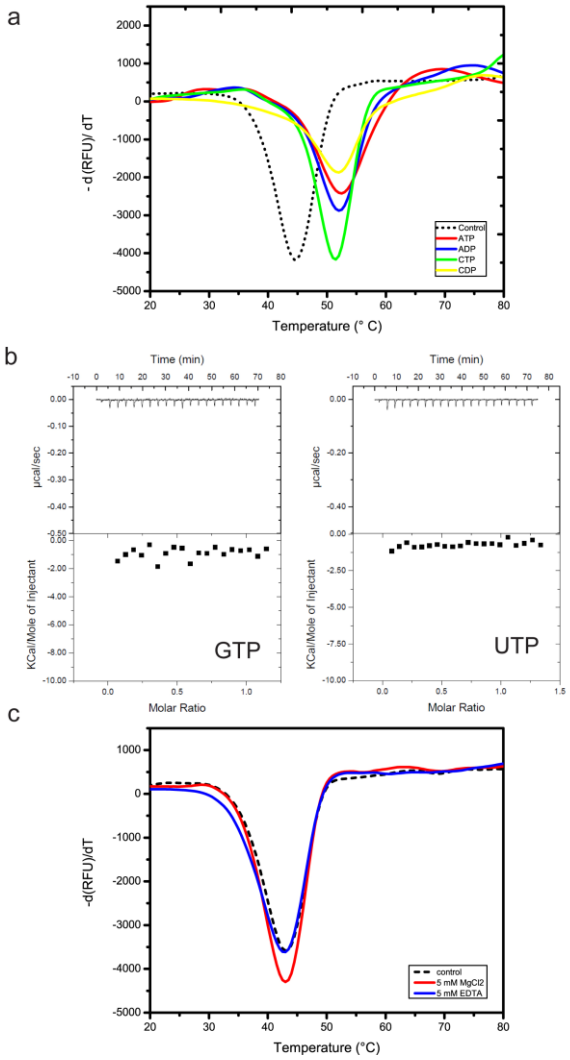
[1]Ludwig-Maximilians-Universität München, Gene Center and Dept. of Biochemistry, Feodor-Lynen-Str. 25, 81377 Munich, Germany

[2]Proteros Biostructures GmbH, Bunsenstraße 7a, 82152 Martinsried, Germany

[3]Center for Integrated Protein Science (CIPSM), Ludwig-Maximilians Universität München, Feodor-Lynen Str. 25, 81377 Munich, Germany

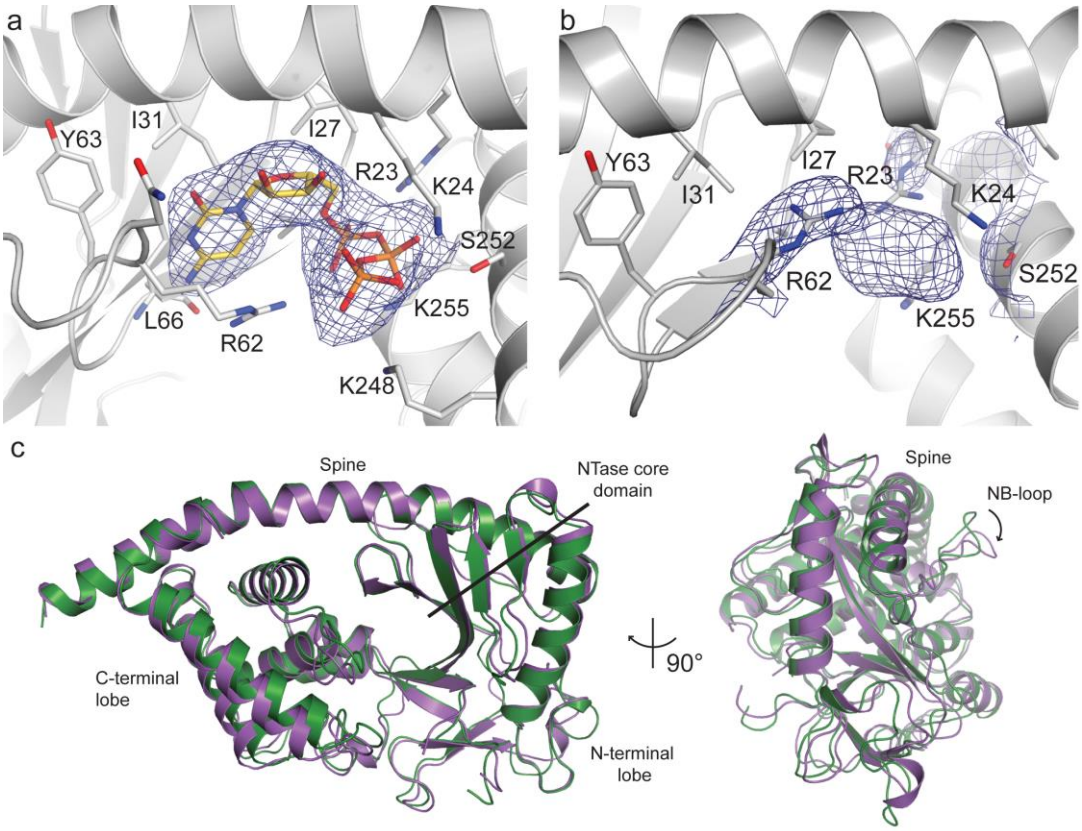


Supplementary figure S1: Patterson self-rotation function, crystal packing and analytical SEC
(a) The Patterson self-rotation function shows five two-fold non-crystallographic symmetry axes at chi=180° and a single five-fold non-crystallographic symmetry axis at chi=72°. (b) MAB21L1 crystal packing forms the decameric assembly (double-pentameric rings) observed in all three crystals, despite the different space groups and unit cell constants. (c) Size-exclusion chromatography of MAB21L1 performed with buffer containing varying NaCl concentrations (150 mM NaCl – red, 500 mM NaCl - purple, 750 mM NaCl - blue, 1M NaCl - green). Purple, blue and green chromatograms overlap. The retention volume of MAB21L1 corresponds to a monomer compared to the gel filtration standard (670 kDa, 158 kDa, 44 kDa, 17 kDa and 1.3 kDa).



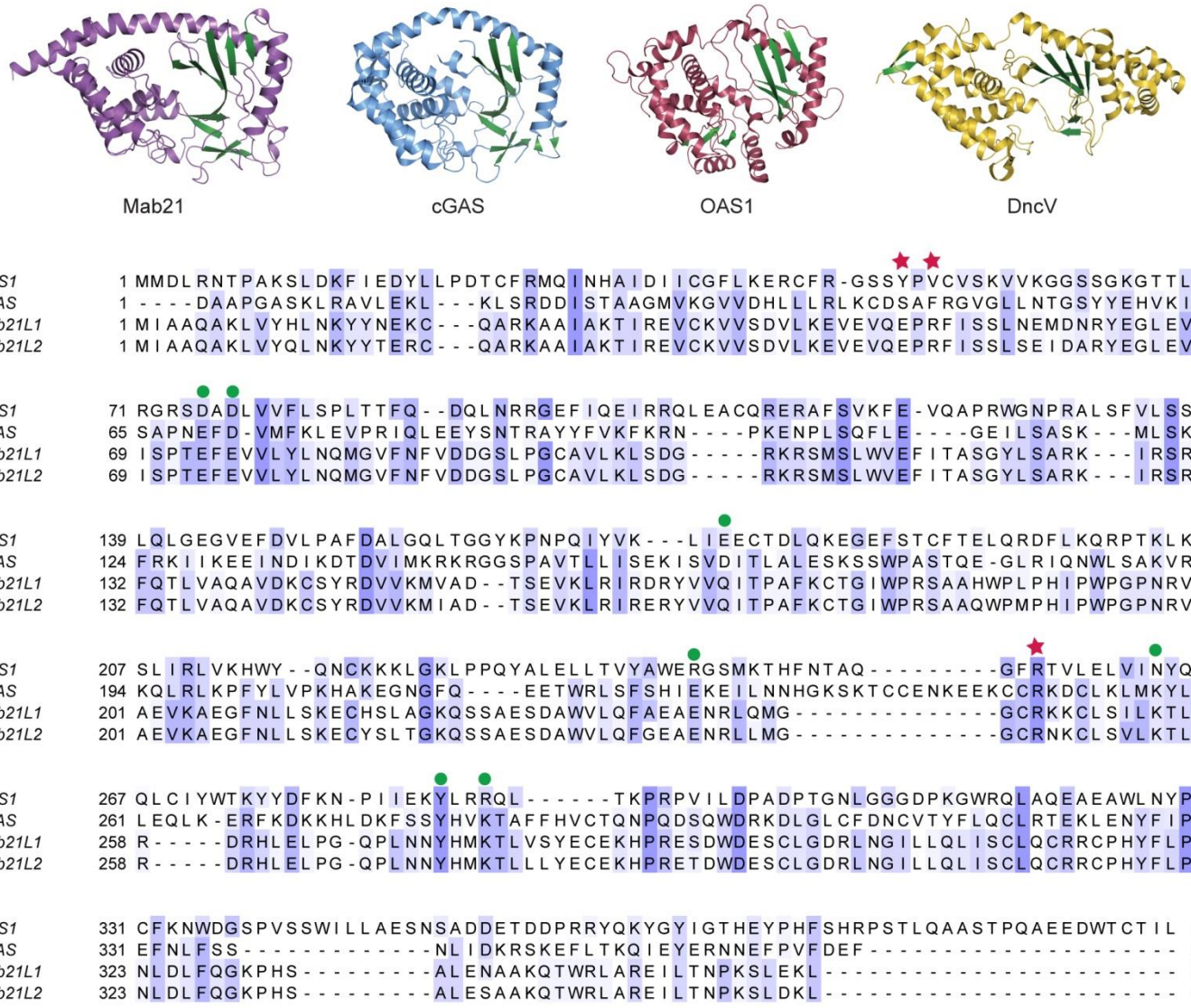
Supplementary figure S2: Fluorescence thermal shift assays and ITC of MAB21L1 with different NTPs and Mg^{2+}

(a) Fluorescence thermal shift assay derivative melt curve plots of MAB21L1 (black, dots), MAB21L1 with 5 mM ATP (red), ADP (blue), CTP (green) and CDP (yellow). (b) Titration of MAB21L1 with GTP and UTP by ITC show no binding to MAB21L1. (c) Fluorescence thermal shift assay derivative melting curve plots of MAB21L1 (black, dots), MAB21L1 in presence of 5 mM $MgCl_2$ (red) and MAB21L1 with 5 mM EDTA (blue).



Supplementary figure S3: Composite omit electron density maps for MAB21L1:CTP and unknown ligand in apoMAB21L1 structure, superposition of apoMAB21L1 and MAB21L1:CTP

(a) Composite omit electron density maps of the ligand binding pocket of MAB21L1:CTP and (b) *apo* MAB21L1 contoured at 1σ. The *apo* MAB21L1 structure shows additional density for an unknown ligand. Key residues interacting with the CTP are depicted and annotated. (c) Front and side views of MAB21L1:CTP (purple) superposed with *apo* MAB21L1 (green) with an RMSD of 0.94Å.



Supplementary figure S4: Structures of the MAB21L1-related proteins cGAS, OAS1 and DncV and a sequence alignment of OAS1, cGAS, MAB21L1 and MAB21L2

Upper panel: crystal structures of MAB21L1 (purple, PDB code 5EOM) and the structurally-related proteins cGAS^{MAB21} (blue, PDB code 4JLX), OAS1 (red, PDB code 4RWQ) and DncV (yellow, PDB code 4XJ1). Lower panel: amino acid sequence alignment of human OAS1, cGAS^{MAB21}, MAB21L1 and MAB21L2 generated with MUSCLE [1] and colored according to the BLOSUM62 conservation score (30%). Green dots mark conserved active site residues involved in metal coordination and conserved residues known to be involved in donor NTP coordination. Red stars denote residues that are mutated in patients with eye malformations in MAB21L2.

1 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797 (2004).

2.3 ATP HYDROLYSIS BY THE VIRAL RNA SENSOR RIG-I PREVENTS UNINTENTIONAL RECOGNITION OF SELF-RNA

Lässig, C., Matheisl, S.*, Sparrer, K. M.*, de Oliveira Mann, C. C.*, Moldt, M., Patel, J. R., Goldeck, M., Hartmann, G., García-Sastre, A., Hornung, V., Conzelmann K.-K., Beckmann R., Hopfner K.-P. (2015) ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *eLife* **4**, e10859

*: equal contribution

In this publication effects of specific structure-derived and patient-related single point mutations in RIG-I's ATPase were analyzed in vitro as well as in human cells. ATP hydrolysis activity coupled to RNA binding is known to be required for SF2 helicases to undergo large conformational changes required for their proper function. So far, it was not clear how RLRs use the driving force of ATP hydrolysis for translocation along the bound RNA in order to distinguish viral from endogenous RNA. Recently, mutations in the ATPase motifs of RIG-I were linked with autoimmune diseases, such as multi-system disorder Singleton-Merten Syndrome (SMS) and Aicardi-Goutières. These mutations lock RIG-I in an active state and lead to constant interferon response. This study showed that mutations disrupting binding of ATP to RIG-I inhibited the ability of the receptor to signal, whereas patient-related mutations still allowed ATP binding but no ATP hydrolysis. ATPase deficient RIG-I constitutively signals through unintentional prolonged binding to endogenous RNA and even co-purified with endogenous RNA from non-infected cells. Here a 60S ribosomal expansion segment was identified as a predominant self-RNA species bound to ATPase deficient RIG-I. In summary, this work suggests that ATP hydrolysis helps to remove/recycle RIG-I from endogenous dsRNA stretches not harboring a 5´-triphosphate, therefore serving as a control strategy for RLRs to overcome nucleic acid self-recognition and limit their signaling to viral non-self RNA.

Author contribution

The author of this thesis contributed to this publication by performing co-immunoprecipitation experiments of RIG-I as well as MDA5 from virus infected and non-infected cells for subsequent analysis of the RNA's immune stimulatory effect. Further, she conducted the thermal shift assays with wild-type RIG-I and mutants.

eLIFE elifesciences.org

RESEARCH ARTICLE

ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA

Charlotte Lässig[1], Sarah Matheisl[1†], Konstantin MJ Sparrer[2†‡], Carina C de Oliveira Mann[1†], Manuela Moldt[1], Jenish R Patel[3,4], Marion Goldeck[5], Gunther Hartmann[5], Adolfo García-Sastre[3,4,6], Veit Hornung[7], Karl-Klaus Conzelmann[2], Roland Beckmann[1,8], Karl-Peter Hopfner[1,8*]

[1]Gene Center, Department of Biochemistry, Ludwig Maximilian University of Munich, Munich, Germany; [2]Max von Pettenkofer-Institute, Gene Center, Ludwig Maximilian University of Munich, Munich, Germany; [3]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, United States; [4]Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, United States; [5]Institute for Clinical Chemistry and Clinical Pharmacology, University Hospital Bonn, University of Bonn, Bonn, Germany; [6]Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, United States; [7]Institute of Molecular Medicine, University Hospital Bonn, University of Bonn, Bonn, Germany; [8]Center for Integrated Protein Science Munich, Munich, Germany

***For correspondence:** hopfner@genzentrum.lmu.de

[†]These authors contributed equally to this work

Present address: [‡] Department of Microbiology and Immunobiology, Harvard Medical School, Boston, United States

Competing interest: See page 16

Funding: See page 16

Received: 14 August 2015
Accepted: 25 November 2015
Published: 26 November 2015

Reviewing editor: Stephen C Kowalczykowski, University of California, Davis, United States

Abstract The cytosolic antiviral innate immune sensor RIG-I distinguishes 5′ tri- or diphosphate containing viral double-stranded (ds) RNA from self-RNA by an incompletely understood mechanism that involves ATP hydrolysis by RIG-I's RNA translocase domain. Recently discovered mutations in ATPase motifs can lead to the multi-system disorder Singleton-Merten Syndrome (SMS) and increased interferon levels, suggesting misregulated signaling by RIG-I. Here we report that SMS mutations phenocopy a mutation that allows ATP binding but prevents hydrolysis. ATPase deficient RIG-I constitutively signals through endogenous RNA and co-purifies with self-RNA even from virus infected cells. Biochemical studies and cryo-electron microscopy identify a 60S ribosomal expansion segment as a dominant self-RNA that is stably bound by ATPase deficient RIG-I. ATP hydrolysis displaces wild-type RIG-I from this self-RNA but not from 5′ triphosphate dsRNA. Our results indicate that ATP-hydrolysis prevents recognition of self-RNA and suggest that SMS mutations lead to unintentional signaling through prolonged RNA binding.
DOI: 10.7554/eLife.10859.001

Introduction

The innate immune system provides a rapid initial reaction to invading pathogens and also stimulates the adaptive immune system (***Iwasaki and Medzhitov, 2015***). Pattern recognition receptors (PRRs) of the innate immune system sense pathogen- or danger-associated molecular patterns (PAMPs or DAMPs) and trigger molecular cascades that together initiate and orchestrate the cellular response through activation of e.g. interferon regulatory factors and nuclear factor κB (***Brubaker et al., 2015***; ***Pandey et al., 2015***; ***Wu and Chen, 2014***).

Retinoic-acid inducible gene I (RIG-I), melanoma differentiation-associated gene 5 (MDA5) and laboratory of physiology and genetics 2 (LGP2) are three structurally related PRRs – denoted RIG-I

eLife digest Living cells produce long, strand-like molecules of RNA that carry the instructions needed to make proteins. Viruses also make use of RNA molecules to hijack an infected cell's protein-production machinery and create new copies of the virus. RNA molecules from viruses have a number of features that distinguish them from a cell's own RNAs, and human cells contain receptors called RLRs that can start an immune response whenever they detect viral RNAs. All of these receptors break down molecules of ATP, a process that releases useable energy. However, so far it is not understood how this activity helps the receptors to distinguish viral RNA from the cell's own RNA molecules (called self-RNA).

Recently, some autoimmune diseases (including Singleton-Merten Syndrome) were linked to mutations in the parts of RLRs that allow the receptors to break down ATP. Now, Lässig et al. have studied the effects of specific mutations in an RLR called RIG-I in human cells. The experiments showed that mutations that disrupt RIG-I's ability to bind to ATP also prevented the receptor from becoming activated. However, mutations linked to Singleton-Merten Syndrome don't stop ATP from binding but instead slow its breakdown; this effectively locks the receptor in an ATP-bound state. Lässig et al. found that similar mutations in RIG-I caused human cells to trigger a constant immune response against the self-RNAs.

Further experiments then suggested that the breakdown of ATP helps to remove RIG-I that has bound to double-stranded sections of self-RNAs. This activity frees the receptor, making it more able to detect double-stranded viral RNAs and preventing unintentional signaling. Lässig et al. also identified a specific double-stranded section of a human RNA that may be recognized by the mutated version of RIG-I in people with Singleton-Merten Syndrome.

The next steps following on from this work are to extend the analysis to also include other RLRs and further explore the underlying mutations within the three-dimensional structures of the receptors and RNA molecules involved.

DOI: 10.7554/eLife.10859.002

like receptors (RLRs) – that recognize cytosolic foreign RNA. RIG-I senses RNA from a broad range of viruses including measles virus and Sendai virus (both paramyxoviridae), Influenza A virus, Japanese encephalitis virus and Hepatitis C virus, whereas MDA5 is activated for example by picornavirus RNA. LGP2 has augmenting and regulatory roles in MDA5 and RIG-I dependent signaling (***Bruns et al., 2014***; ***Satoh et al., 2010***; ***Sparrer and Gack, 2015***).

RIG-I preferentially detects base-paired double-stranded RNA (dsRNA) ends containing either 5′ triphosphate (ppp) or 5′ diphosphate (pp) moieties (***Goubau et al., 2014***; ***Hornung et al., 2006***; ***Pichlmair et al., 2006***; ***Schlee et al., 2009***; ***Schmidt et al., 2009***) and not 2′ OH methylated at the first 5′ terminal nucleotide (***Schuberth-Wagner et al., 2015***). ppp-dsRNA arises, for example, at panhandle structures of influenza virus nucleocapsids, or during measles or Sendai virus transcription (***Liu et al., 2015***; ***Weber et al., 2013***). 5′ diphosphates are found on genomic RNA of reoviruses (***Banerjee and Shatkin, 1971***). RIG-I can also detect poly-U/UC-rich dsRNA (***Schnell et al., 2012***). Ligands of MDA5 are less well characterized but include dsRNA longer than 1000 bp (***Kato et al., 2008***), higher-order dsRNA structures (***Pichlmair et al., 2009***), or AU-rich RNA (***Runge et al., 2014***).

RLRs are members of the superfamily II (SF2) of ATPases, helicases or nucleic acid translocases. RIG-I and MDA5 consist of two N-terminal tandem caspase activation and recruitment domains (2CARD), a central ATPase/translocase domain and a C-terminal regulatory domain (RD). LGP2 lacks the 2CARD module but otherwise has a similar domain architecture. Binding of RNA induces a conformational change in RIG-I. If activated, the RD binds the ppp- or pp-dsRNA end, while the SF2 domain interacts with the adjacent RNA duplex and forms an active ATPase site (***Civril et al., 2011***). In this conformation, the 2CARD module is sterically displaced from its auto-inhibited state (***Jiang et al., 2011***; ***Kowalinski et al., 2011***; ***Luo et al., 2011***) and can be K63-linked poly-ubiquitinated (***Gack et al., 2007***). Multiple Ub-2CARD complexes assemble to form a nucleation site for the polymerization of mitochondrial antiviral-signaling adaptor protein (MAVS) into long helical filaments (***Hou et al., 2011***; ***Wu et al., 2014***; ***Xu et al., 2014***). Instead of recognizing terminal structures like

RIG-I, MDA5 cooperatively polymerizes along dsRNA (**Berke and Modis, 2012**), which is suggested to trigger MAVS polymerization.

The SF2 ATPase domain plays a critical part in RIG-I activation, although the role of the ATPase activity is still debated. Mutation of the seven SF2 "helicase" motifs resulted in RLRs that are either inactive or signal constitutively (**Bamming and Horvath, 2009**; **Louber et al., 2015**). On the other hand, overexpression of the 2CARD module alone is sufficient for signaling (**Yoneyama et al., 2004**). Further studies revealed that the SF2 domain is an ATP-dependent dsRNA translocase (**Myong et al., 2009**) that can help enhance signaling by loading multiple RIG-I on dsRNA (**Patel et al., 2013**) and may execute anti-viral "effector" functions through displacement of viral proteins (**Yao et al., 2015**). Finally, RIG-I ATPase activity promotes recycling of RIG-I:dsRNA complexes in vitro, suggesting a kinetic discrimination between self and non-self RNA (**Anchisi et al., 2015**; **Louber et al., 2015**).

Several autoimmune diseases, including the Aicardi-Goutières and Singleton-Merten syndromes (SMS), were linked to single amino acid mutations in the SF2 domains of MDA5 and RIG-I (**Funabiki et al., 2014**; **Jang et al., 2015**; **Rice et al., 2014**; **Rutsch et al., 2015**). Two point mutations within the Walker A (motif I) or Walker B (motif II) of RIG-I are linked to atypical SMS and functional studies indicated constitutive RIG-I activation (**Jang et al., 2015**). Thus, these mutations have been described as a gain of function, which is puzzling considering previous mutations in motif I led to loss of RIG-I function, while mutations in motif II led to either gain or loss of function, depending on the type of mutation (**Bamming and Horvath, 2009**; **Louber et al., 2015**).

In order to clarify the role of RIG-I's ATPase in antiviral signaling and RLR associated human diseases, we engineered structure-derived and patient-identified mutations into RIG-I and tested the resulting proteins in different types of cell-based and in vitro analyses. Collectively, we find that SMS mutations phenocopy the structure-derived E373Q mutation in motif II, which is designed to trap RIG-I in an ATP-bound state. Freezing this state results in a dramatic autoimmune response because the enzyme binds self-RNA and signals. An unexpected, strongly enriched self-RNA is the ribosomal large subunit, which contains large, dsRNA expansion segments. Collectively, our results suggest that a biomedical and functional critical role of RIG-I's ATPase is to prevent spontaneous and unintended activation by self-RNA. Thus, the SF2 translocase likely increases the sensitivity of the system by reducing background signaling. Furthermore, our studies suggest that in SMS, RIG-I is trapped in an ATP-bound state and signals through self-ligands.

Results

Prevention of ATP hydrolysis in RIG-I leads to a constitutive activation of the interferon-β promoter by recognition of self-RNA

To address the roles of ATP binding and hydrolysis by the SF2 domain of RIG-I, we studied RIG-I variants containing structure-based mutations designed to i) prevent ATP binding and formation of a functional ATP-bound complex, ii) allow ATP binding and ATP-induced conformational changes but prevent ATP hydrolysis, or iii) disable interaction of the RNA with either the 1A or 2A domain of SF2 (**Figure 1A, B**). The structure of RIG-I in complex with RNA and ADP·BeF_x served as guide for these mutations ([**Jiang et al., 2011**], PDB code 3TMI, **Figure 1B**).

In order to dissect the influence of these mutations on the ability of RIG-I to elicit downstream signaling, we used an interferon-β (IFNβ) promoter activity assay carried out in HEK 293T RIG-I KO cells (**Figure 1—figure supplement 1A,B**). Overexpressed wild-type RIG-I (wtRIG-I) is able to induce a slight activation of the IFNβ promoter, which can be further amplified by stimulation with Sendai virus defective interfering particles (SeV DIs) (**Figure 1C**). The 2CARD module (RIG-I 1-229) induced a strong activation in both non-infected and SeV DI-stimulated cells and is crucial since constructs lacking these domains (Δ2CARD, RIG-I 230-925) cannot conduct any downstream signaling. RIG-I K270I, carrying a mutation in the motif I lysine that reduces ATP binding (**Rozen et al., 1989**), signaled in neither uninfected nor SeV DIs stimulated cells, consistent with previous studies. Remarkably, the E373Q substitution in motif II had a strikingly different effect. RIG E373Q, which has a stabilized ATP-bound state by slowed-down ATP hydrolysis, strongly signaled in both non-infected and SeV DIs stimulated cells. Western blots validated correct expression of all mutants (**Figure 1—figure supplement 1C**).

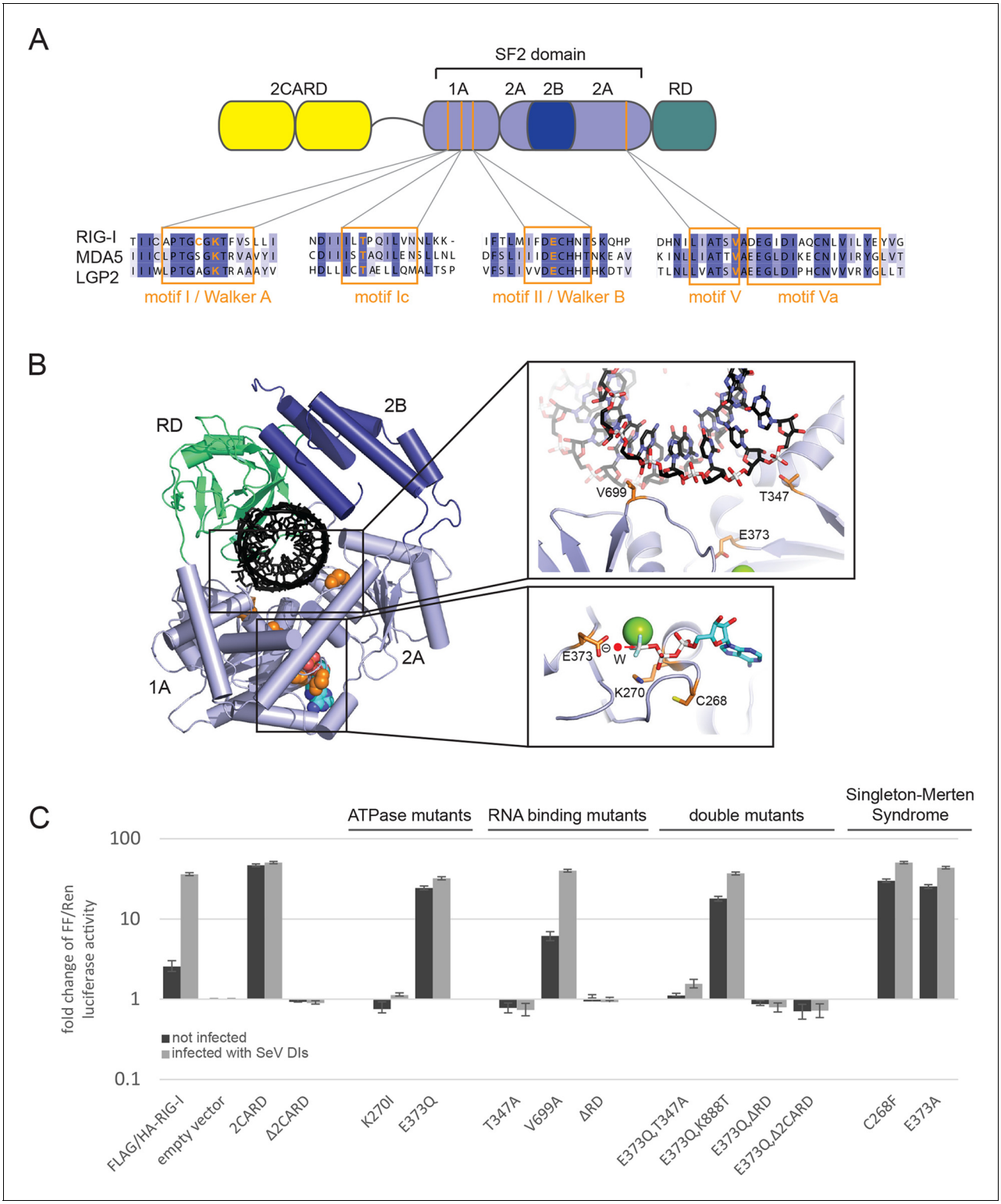


Figure 1. Cellular studies of RIG-I ATPase mutants in infected or non-infected cells. (**A**) Location of amino acid substitutions of RIG-I SF2 domain variants used in this study (orange lines) within different RLR helicase motifs (orange squares). (**B**) Single amino acid substitutions (orange) within the RIG-I 3D structure (PDB: 3TMI). (**C**) Fold change of interferon-β (IFNβ) promoter driven luciferase activity in uninfected HEK 293T RIG-I KO cells or in cells challenged with Sendai virus defective interfering particles (SeV DIs). Cells were co-transfected with RIG-I expression vectors and p125-luc/ pCMV-RL reporter plasmids, and infected with SeV DIs 6 hr post transfection. Firefly (FF) luciferase activities were determined in respect to Renilla (Ren) luciferase activities 24 hpi. All ratios were normalized to the empty vector control. n=3–12, error bars represent mean values ± standard deviation.

Figure 1 continued on next page

To rule out a "constitutive" active conformation of RIG-I E373Q due to an exposed 2CARD module (e.g. from an unfolded SF2) we performed small angle X-ray scattering with purified wtRIG-I and RIG-I E373Q demonstrating that both proteins have the same solution structure (***Figure 1—figure supplement 2A, B***). In addition, thermal unfolding assays show that the E373Q mutation does not destabilize RIG-I (***Figure 1—figure supplement 2C***). Finally, RIG-I Δ2CARD,E373Q has a dominant negative effect on signaling by RIG-I E373Q (***Figure 1***, ***Figure 1—figure supplement 2D, E***). Taken together, these data show that RIG-I E373Q is neither destabilized nor constitutively active, suggesting it needs productive RNA interactions.

To test whether E373Q signals in non-infected (and perhaps also infected cells) because of interaction with self-RNA, we additionally introduced mutations in various RNA binding sites, in particular a ΔRD variant (RIG-I 1-798) and mutations in two RNA-interacting residues in domains 1A (T347A) and 2A (V699A) of SF2. The single mutation RIG-I T347A did not signal in either infected or non-infected cells, showing that the interaction of RNA with this specific amino acid is critical for signaling (***Figure 1C***). Interestingly, we find that the single mutation V699A slightly increases the signaling activity of RIG-I in non-infected cells (***Figure 1C***), which could be explained by a putative reduction of translocation activity instead of a prevention of RNA binding to SF2 (see discussion). Finally, deletion of the regulatory domain (ΔRD) inactivates signaling in both infected and non-infected cells as previously observed (***Cui et al., 2008***). As expected, both combination mutants RIG-I E373Q,T347A and RIG-I E373Q,ΔRD failed to signal in both SeV DIs infected and non-infected cells. These data show that the increased immunostimulatory effect of E373Q requires a productive RNA interaction of SF2 and RD.

Since RD is also required for the displacement of the 2CARD module from SF2, we additionally analyzed a point mutation in RD. K888 mediates triphosphate binding in RD and mutations in this residue inactivate recognition of viral RNA (***Cui et al., 2008***; ***Wang et al., 2010***). Of note RIG-I E373Q,K888T is still constitutively active in non-infected cells. This effect indicates that the increased signaling capacity on endogenous RNA is independent from the ppp-dsRNA or pp-dsRNA epitopes that RIG-I recognizes on viral RNA via the RD.

Finally, we addressed the effect of the Singleton-Merten mutations C268F and E373A. E373A is at the same position as our structure-derived E373Q mutant. Consistent with this, we observed that this substitution leads to a constitutive activation of the IFNβ promoter (***Jang et al., 2015***) (***Figure 1C***). Interestingly, although C268 is located in motif I, it also leads to constitutive signaling, whereas motif I mutation of K270 (which coordinates the β-phosphate of ATP) blocks ATP binding and renders RIG-I inactive. Thus, mutation of the non-ATP binding C268 in motif I appears to phenocopy a mutation that prevents ATP hydrolysis.

In summary, our studies show that signaling of RIG-I requires both ATP and RNA binding. ATP hydrolysis, on the other hand, appears to be critical to dissolve the signaling state and to prevent activation of RIG-I by self-RNA.

RIG-I ATP hydrolysis defective mutant E373Q shows increased interaction with ribosomal RNA

We hypothesized that E373Q traps RIG-I in an ATP bound high affinity conformation that is activated already by self-RNA. To test this idea, we immunoprecipitated RIG-I and its mutants from non-infected HEK 293T RIG-I KO cells or cells infected with measles or Sendai virus and analyzed the co-purified RNA molecules. Regardless of whether co-purified from infected or non-infected cells, the amount of RNA recovered from RIG-I E373Q was about 3 times higher than that from RIG-I (***Figure 2A***). Similarly increased amounts of RNA co-purified with the SMS mutants C268F and E373A from uninfected cells, reflecting the same altered RNA binding properties as in RIG-I E373Q (***Figure 2—figure supplement 1A***).

When analyzed on a Bioanalyzer RNA chip or on agarose gels, we found that the increased amount of RNA is to a large extent due to the presence of 28S rRNA, while 18S rRNA remains unaltered (***Figure 2B***). Control analysis of the total RNA content ruled out an alteration of ribosome subunit ratio in RIG-I E373Q transfected cells (***Figure 2C***). Both increased amount of RNA and specific enrichment of 28S rRNA were also observed for the equivalent MDA5 E444Q Walker B mutant (***Figure 2—figure supplement 1B, C***).

In order to determine the immunostimulatory potential of the RNA co-purified from virus-infected cells, we back-transfected the RNA into HEK 293T ISRE-FF/RFP reporter cells (which contain

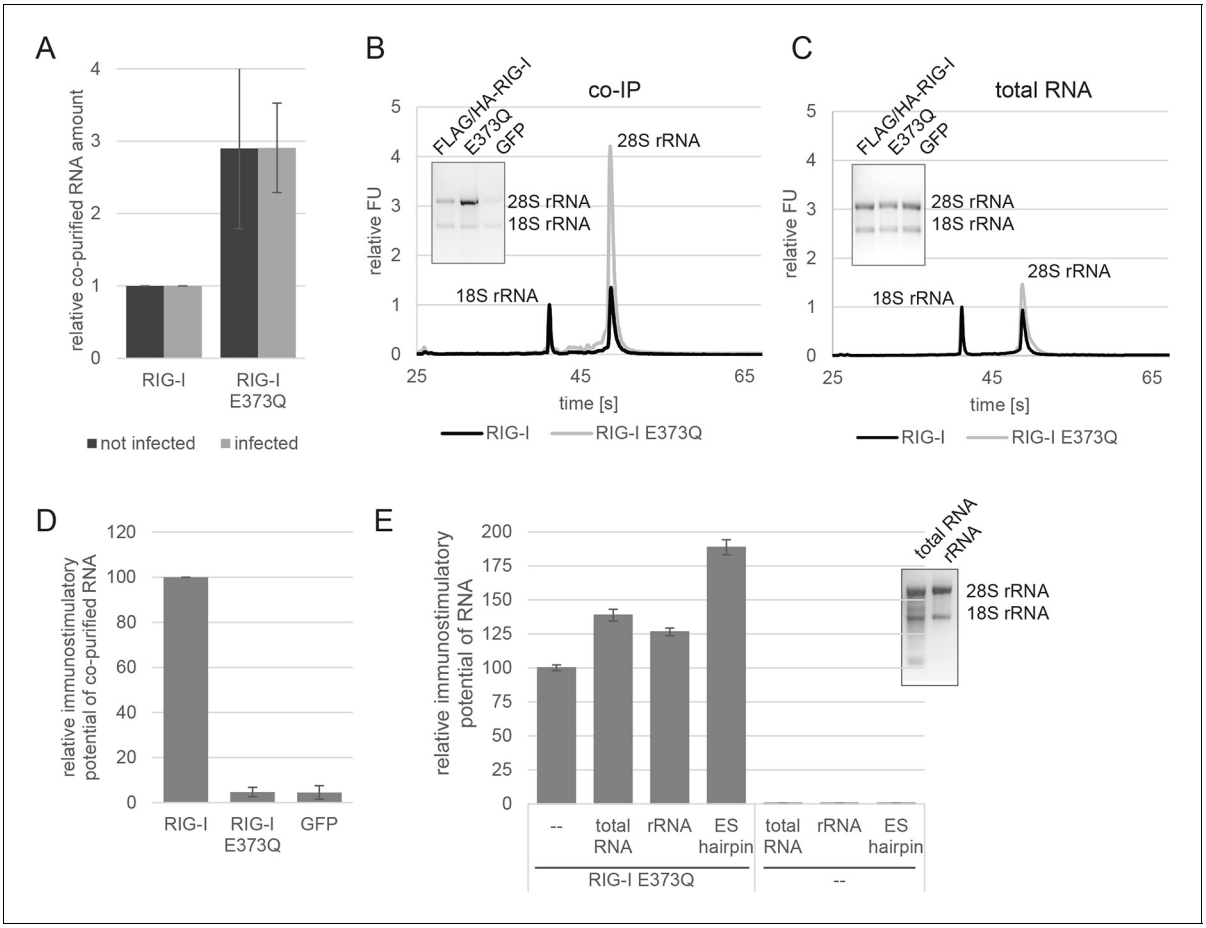


Figure 2. RIG-I ATP hydrolysis defective mutant E373Q recognizes the 60S ribosomal subunit in vivo. (**A**) Relative RNA amount co-purified with overexpressed RIG-I or RIG-I E373Q from virus infected or non-infected HEK 293T RIG-I KO cells. n=4 (infected) or n=10 (non-infected), error bars represent mean values ± standard deviation. (**B**) Bioanalyzer evaluation and agarose gel separation of RNA co-purified with overexpressed RIG-I or RIG-I E373Q from non-infected HEK 293T RIG-I KO cells. Curves are normalized in respect to 18S rRNA peaks. (**C**) Bioanalyzer evaluation and agarose gel separation of total RNA content of non-infected HEK 293T RIG-I KO cells overexpressing RIG-I or RIG-I E373Q. Curves were normalized as in panel B. (**D**) Immunostimulatory potential of co-purified RNA from RIG-I, RIG-I E373Q or GFP overexpressed in measles virus (MeV), MeV-Cko-ATU-Cs or Sendai virus Cantell (SeV) infected HEK 293T RIG-I KO cells. RNA was back-transfected into HEK 293T ISRE-FF/RFP cells together with pTK-RL transfection control. Firefly luciferase (FF) activities were determined 24 hr after transfection in respect to Renilla luciferase (Ren) activity and were normalized to the immunostimulatory potential of RIG-I associated RNA. n=4, error bars represent mean values ± standard deviation. (**E**) Immunostimulatory potential of endogenous RNA in cells overexpressing RIG-I E373Q. RNA was co-transfected into HEK 293T RIG-I KO cells together with a RIG-I E373Q expression vector and p125-luc/ pCMV-RL reporter plasmids. FF luciferase activities were determined in respect to Ren luciferase activities 24 hr after transfection. All ratios are normalized to the RIG-I E373Q control without RNA stimulation. Purified RNA was in addition analyzed on agarose gels. n=3, error bars represent mean values ± standard deviation.

DOI: 10.7554/eLife.10859.006

The following figure supplements are available for figure 2:

Figure supplement 1. Analysis of RNA co-purified with RIG-I SMS or MDA5 variants.

DOI: 10.7554/eLife.10859.007

Figure supplement 2. Assay for defining the immunostimulatory potential of different RNAs.

DOI: 10.7554/eLife.10859.008

Figure supplement 3. Immunostimulatory potential of co-purified RNA from Sendai virus Cantell (SeV) infected cells.

DOI: 10.7554/eLife.10859.009

endogenous RIG-I, see ***Figure 2—figure supplement 2A***). RNA co-purified with wtRIG-I and RIG-I lacking the 2CARD module induced an immune response in these cells (***Figure 2—figure supplement 3A***). RNA co-purified with RIG-I K270I (ATP binding deficient) and V699A (putative translocation deficient) was also able to stimulate the ISRE reporter in an amount comparable to wtRIG-I, indicating no altered RNA binding properties in these mutants under virus infected conditions. In

contrast, RNA that co-purified with the RNA-binding deficient RIG-I T347A (mutation in SF2 domain), RIG-I K858E (mutation in RD domain that reduces triphosphate recognition) or RIG-I ΔRD poorly stimulated the ISRE promoter and probably represents background RNA (***Figure 2—figure supplement 3A***). These data suggest that RIG-I recognizes immunostimulatory RNA via the SF2 and RD domains, but does not require ATP binding for this process. ATP binding is necessary, however, because RIG-I K270I expression alone does not stimulate the IFNβ promoter (compare with ***Figure 1C***). Interestingly, RNA co-purified with RIG-I E373Q failed to induce reporter gene expression (***Figure 2D***, ***Figure 2—figure supplement 3A***). Thus, despite the observation that RIG-I E373Q co-purifies with approximately threefold more RNA than wtRIG-I from infected cells, the co-purified RNA is not immunostimulatory in a wtRIG-I background. However, cells that transiently express RIG-I E373Q can be further stimulated by transfection of total RNA extracts and purified ribosomal RNA (***Figure 2E***), suggesting that ribosomal RNA can activate RIG-I E373Q. Cells lacking wtRIG-I or RIG-I E373Q on the other hand do not respond to those RNAs. We conclude that host-RNA, which does not activate wtRIG-I, can apparently compete with viral RNA for RIG-I E373Q.

In order to verify a higher affinity of the RIG-I ATP hydrolysis defective mutant towards ribosomal RNA, we purified full-length human RIG-I and RIG-I E373Q, as well as human 80S ribosomes, and tested for a direct interaction. We confirmed that while both RIG-I E373Q and the wild-type protein are able to bind ATP, only wtRIG-I can hydrolyze ATP (***Figure 3A, B***). We subsequently conducted sedimentation assays via ultra-centrifugation of sucrose cushions loaded with 80S ribosomes that have been pre-incubated with wtRIG-I or RIG-I E373Q in presence or absence of ATP or the non-hydrolysable ATP analogue ADP·BeF_3. In presence of ATP a minor binding of wtRIG-I to the ribosome could be observed, whereas RIG-I E373Q bound in a near stoichiometric manner. In absence of ATP or in presence of ADP·BeF_3 binding of wtRIG-I was greatly enhanced and showed similar levels compared to RIG-I E373Q (***Figure 3C***).

We next analyzed RIG-I E373Q:80S ribosome complexes by cryo-electron microscopy and single particle 3D reconstruction (***Figure 3D***). The average resolution was estimated to be 17.7 Å based on the Fourier shell correlation cut-off criterion at 0.5. When compared with the reconstruction of the human 80S ribosome alone (***Figure 3E***), the ribosome:RIG-I E373Q complex revealed an additional density located at rRNA expansion segment (ES) 7L, which is located at the back of the large ribosomal subunit. Calculation of a statistical difference map between the two reconstructions confirmed that this distinct region contained significant additional density (***Figure 3F***). Human ribosomes contain several long, G:C rich, base-paired RNA expansion segments forming large tentacle-like hairpin structures of substantial double-stranded nature (***Anger et al., 2013***). A large part of the double-stranded RNA in these segments is not covered by ribosomal proteins and accessible for cytosolic proteins. The crystal structure of ADP·BeF_x-bound RIG-I △2CARD:RNA complex ((***Jiang et al., 2011***), PDB code 3TMI) fits well into the density observed at ES7L and is located at the root of the solvent exposed portion of helix A of ES7L that contains a contiguous stretch of seven G:C/C:G base pairs (***Figure 3G***).

In summary, we conclude that stabilizing the ATP-bound state of RIG-I induces a conformation where RIG-I binds to ribosomes, presumably at exposed dsRNA expansion segments.

Specificity of RIG-I towards double-stranded RNA is increased in presence of ATP

To further evaluate the role of ATP binding and hydrolysis of RIG-I we performed electrophoretic mobility shift assays (EMSAs), fluorescence anisotropy experiments and ATP hydrolysis assays in presence and absence of ATP or ADP·BeF_3 with different RNAs. These RNAs mimic different types of endogenous or viral RNAs and help dissecting contributions of RD's binding to the RNA end and SF2's binding to the stem. In addition to a 24mer or 12mer blunt-ended dsRNA or ppp-dsRNA (***Goldeck et al., 2014***), we also used a 60 nucleotide hairpin RNA (denoted as ES hairpin) derived from the ribosomal expansion segment ES7L, which contains several bulges and a non-pairing end (***Figure 4—figure supplement 1A***). The hairpin at one end and the added Y-structure at the other end are used to minimize RNA end binding by RIG-I's RD because RD has a high affinity for blunt RNA ends.

RIG-I and RIG-I E373Q bound to the 24mer blunt ended dsRNA with a slightly higher affinity in presence of ATP or ADP·BeF_3 than in its absence (***Figure 4A***), suggesting that ATP binding to the SF2 domain positively contributes to the overall affinity in addition to RD. A similar result was

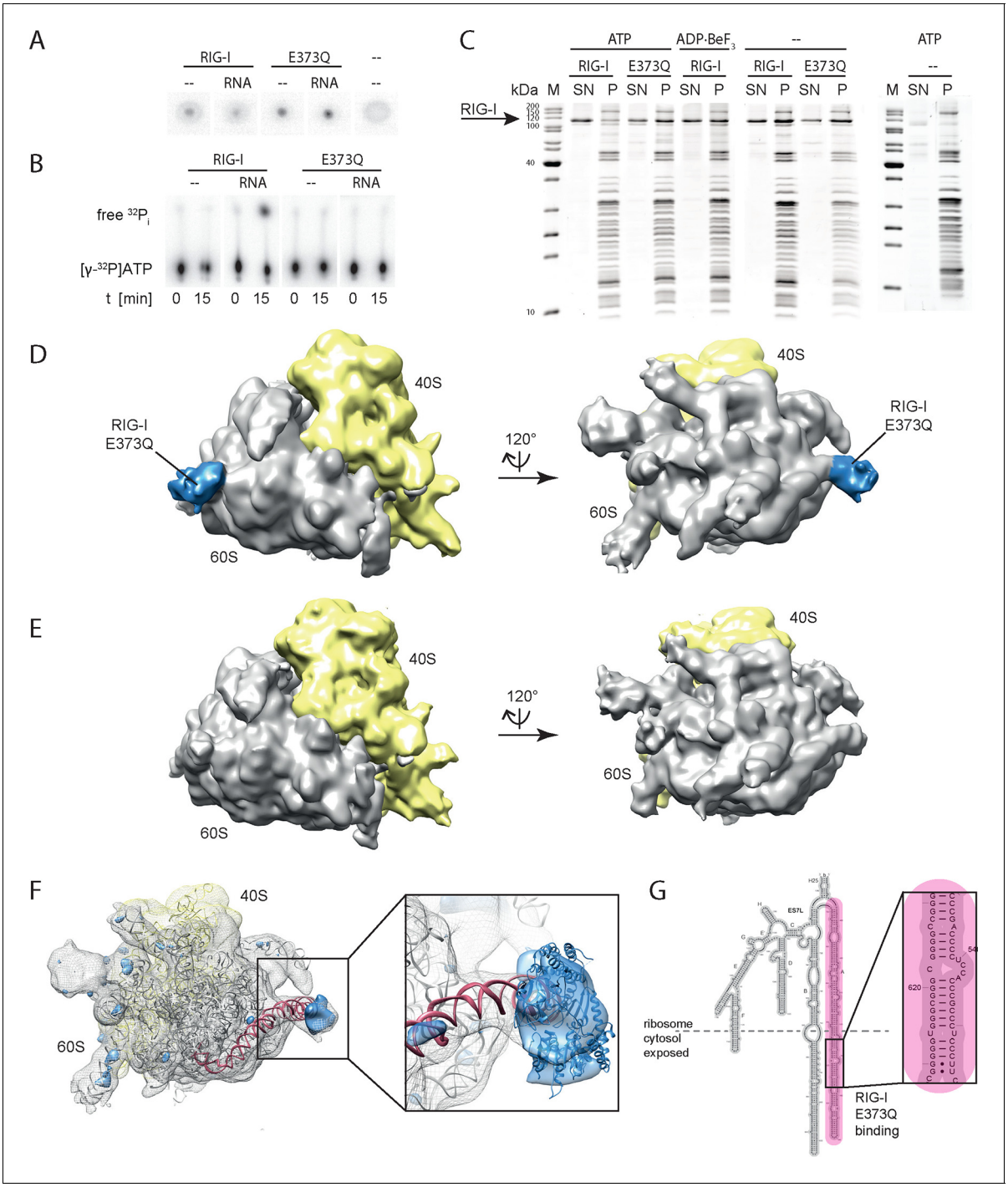


Figure 3. RIG-I ATP hydrolysis defective mutant E373Q recognizes the 60S ribosomal subunit in vitro. (**A**) DRaCALA ATP binding assay of RIG-I or RIG-I E373Q in presence or absence of RNA. (**B**) ATP hydrolysis assay of RIG-I or RIG-I E373Q in presence and absence of RNA. (**C**) Binding studies of human 80S ribosomes with RIG-I or RIG-I E373Q in presence or absence of ATP or ADP·BeF_3. Pre-formed complexes were separated on sucrose cushions via ultracentrifugation and pellet (P) as well as supernatant (SN) fractions were analyzed by SDS-PAGE. (**D**) Side views of a cryo-EM reconstruction of RIG-I E373Q (blue) bound to the human 80S ribosome (yellow: 40S subunit, gray: 60S subunit). Data was low pass-filtered at 15 Å. (**E**) Side views of a cryo-EM reconstruction of the human 80S ribosome without prior RIG-I E373Q incubation. Data filtering and color coding as in panel D. (**F**) Statistical difference map (left, σ = 2) of cryo-EM reconstructions in panels D and E reveals a significant additional density at expansion segment 7L A (ES7L-A, pink) into which RIG-I (PDB 3TMI) can be fitted (right, σ = 1.51). (**G**) Secondary structure map of the 28S rRNA ES7L (derived from (***Anger et al., 2013***) and zoom into RIG-I E373Q binding area. ES7L-A is indicated in pink (as in panel **F**).

DOI: 10.7554/eLife.10859.010

obtained when we used a 12mer dsRNA in fluorescence anisotropy experiments in order to further dissect the influence of different RNA ends (***Figure 4B***). Interestingly, the positive effect of ATP was not observed when we used the corresponding ppp-dsRNA 12mer (***Figure 4C***), most likely because the RD dominates RNA binding under these conditions. Thus, it is plausible that RIG-I dissociates from unphosphorylated RNA termini with an increased rate after ATP hydrolysis than from triphosphorylated termini.

We next tested the role of ATP on binding of wtRIG-I, RIG-I E373Q, RIG-I T347A,E373Q and the SMS variant RIG-I C268F to the ES hairpin RNA mimicking the base of the ribosomal ES7L. In presence of ATP we observed moderately increased binding of RIG-I E373Q and of RIG-I C268F to this hairpin, however wtRIG-I displayed a strikingly opposing effect (***Figure 4D***, ***Figure 4—figure supplement 1C***). For this RNA, ATP reduced rather than increased the affinity of wtRIG-I. The addition of ADP·BeF_3 to RIG-I could reconstitute the high affinity state of RIG-I E373Q. The RIG-I T347A,E373Q double mutant, on the other hand, showed binding affinities similar to RIG-I in presence of ATP, probably caused by residual binding of RD (***Figure 4D***).

Consistent with this, the ES hairpin RNA could induce signaling in RIG-I E373Q transfected HEK 293T RIG-I KO cells (***Figure 2F***) and could also stimulate the ATPase activity of RIG-I △2CARD, and to a lesser extent wtRIG-I (which is auto-inhibited by the 2CARD module) (***Figure 5A***, ***Figure 5—figure supplement 1A***). A comparable stimulatory effect on the ATPase activity of RIG-I could also be detected with whole human ribosomes (***Figure 5A***). Control assays with the ATP hydrolysis defective mutants RIG-I E373Q and RIG-I T347A,E373Q confirmed the lacking ability of those proteins to hydrolyze ATP even in the presence of triphosphorylated RNA (***Figure 5A***, ***Figure 5—figure supplement 1B***).

In summary, our results show that ATP hydrolysis leads to a moderately increased binding of RNA containing base-paired ends, but decreased binding of RNA lacking base-paired ends. These in vitro data are also consistent with our co-immunopurification studies of RNA from cells, where we observed that the ATP hydrolysis deficient RIG-I E373Q mutant co-purified with increased amounts of endogenous RNA.

Discussion

Here we show that mutations that slow down or inhibit RIG-I's ATPase lead to an increased interaction of RIG-I with endogenous RNA, including double-stranded RNA expansion segments of the human large ribosomal subunit. Our results suggest that RIG-I's ATPase confers specificity to viral RNA by preventing signaling through the abundant background of self-RNA and provide a molecular framework for understanding the pathology of atypical Singleton-Merton syndrome.

Recently, several autoimmune diseases, including the Aicardi-Goutières and Singleton-Merten syndromes, have been linked to RLRs through whole exome sequencing, which discovered single amino acid mutations that are mostly found within the ATPase domain of RLRs (***Jang et al., 2015***; ***Rice et al., 2014***; ***Rutsch et al., 2015***). Increased interferon levels suggest that an increased activation of MDA5 or RIG-I underlies the molecular pathology of these diseases. Indeed we find that not only E373Q, consistent with recent results, leads to an increased activation of RIG-I in non-infected cells, but also the SMS mutations E373A and C268F (***Jang et al., 2015***) (***Figure 1C***). While this could have been expected for E373A, because of its similarity to E373Q, the increased immunostimulatory effect of C268F in motif I comes as a surprise. Prior mutations in motif I studied by others and us led to an inactivation of RIG-I, rather than constitutive activation. The precise structural reason for the increased signaling of C268F needs to be addressed in future studies, but our co-immunoprecipitation and in vitro binding assay results suggest that this mutation may also lock RIG-I in an RNA-bound, active conformation (***Figure 2—figure supplement 1A***, ***Figure 4—figure supplement 1C***).

Mutational and biochemical analyses previously suggested a kinetic model for RIG-I's specificity towards viral RNA, where the ATP-dependent recycling helps to discriminate ppp-dsRNA from endogenous RNA (***Anchisi et al., 2015***; ***Louber et al., 2015***; ***Runge et al., 2014***) (***Figure 6A***). Our studies show that, in case of base-paired triphosphate containing RNA ends, the RIG-I RD dominates binding. Although RIG-I's ATPase is very active, we do not see a strong effect of ATP on the affinity for the RNA (***Figure 4C***, ***Figure 5A***). ATP hydrolysis may under the assayed conditions not efficiently displace RIG-I from ppp-dsRNA because RD might prevent full dissociation even after ATP-hydrolysis displaced SF2. Importantly, ATP reduces the affinity towards self-RNA containing a duplex region

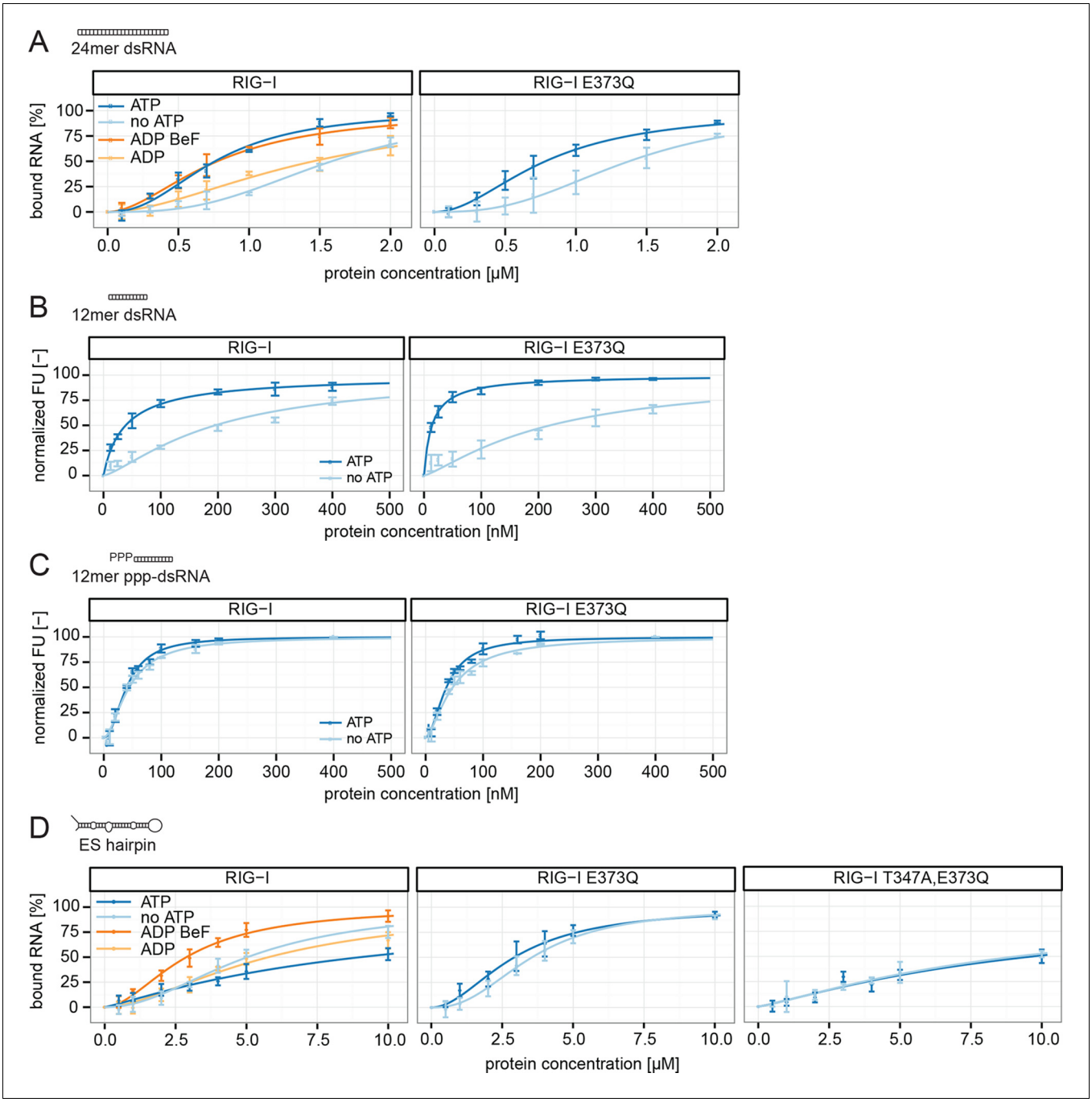


Figure 4. RIG-I's ATP hydrolysis enhances RNA end recognition and removes RIG-I from RNA stems. (**A**) Quantification of electrophoretic mobility shift assays of RIG-I or RIG-I E373Q incubated with 24mer dsRNA in presence or absence of ATP, ADP or ADP·BeF_3 (compare with ***Figure 4—Figure supplement 1B***). (**B**) Fluorescence anisotropy changes measured by titrating RIG-I or RIG-I E373Q in presence or absence of ATP into solutions containing fluorescently labeled 12mer dsRNA. (**C**) Fluorescence anisotropy changes measured by titrating RIG-I or RIG-I E373Q in presence or absence of ATP into solutions containing fluorescently labeled 12mer ppp-dsRNA. (**D**) Quantification of electrophoretic mobility shift assays of RIG-I, RIG-I E373Q or RIG-I T347A, E373Q incubated with an RNA hairpin derived from helix A of the human ribosome expansion segment 7L (ES hairpin) in presence or absence of ATP, ADP or ADP·BeF_3 (compare with ***Figure 4—Figure supplement 1C***). All binding curves were fitted using the LL.2 function of the R drc package (***Cedergreen et al., 2005***). n=3-6, error bars represent mean values ± standard deviation.

DOI: 10.7554/eLife.10859.011

The following figure supplement is available for figure 4:

Figure supplement 1. Design of the ribosomal expansion segment derived hairpin RNA, EMSA raw figures and control experiments with RIG-I C268F SMS mutant.

DOI: 10.7554/eLife.10859.012

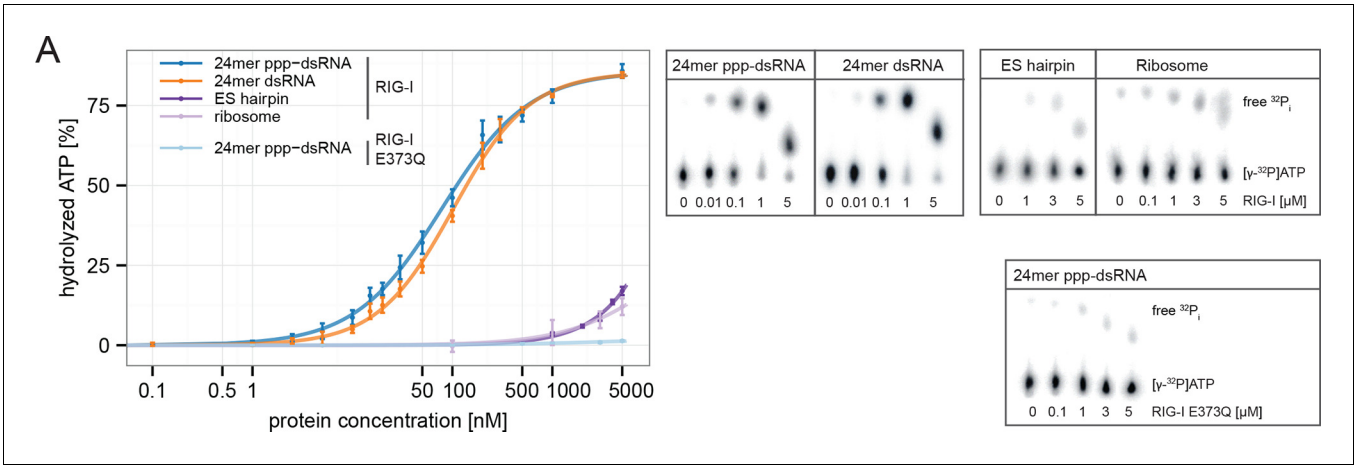


Figure 5. RIG-I's ATPase activity correlates with its RNA binding affinity. (**A**) Quantification of hydrolyzed [γ-^{32}P]ATP by RIG-I or RIG-I E373Q in presence of different RNA substrates. Reactions were allowed to proceed for 20 min at 37 °C and free phosphate was separated from ATP via thin layer chromatography. Spots corresponding to labeled ATP and labeled P_i were quantified using ImageJ. All curves were fitted using the LL.2 function of the R drc package. n=3, error bars represent mean values ± standard deviation.

DOI: 10.7554/eLife.10859.013

The following figure supplement is available for figure 5:

Figure supplement 1. RIG-I's 2CARD module reduces the ATP hydrolysis activity.

DOI: 10.7554/eLife.10859.014

but not a "proper" ppp-dsRNA end (***Figure 4D***). Thus, if RD is unable to tether RIG-I to ppp-dsRNA ends the ATPase could rapidly remove RIG-I from RNA duplex regions via its translocase and therefore prevents an autoimmune response towards self-RNA (***Figure 6B***). Our cellular studies are consistent with this biochemical observation, because a point mutation in K888, a residue that is critical for recognizing ppp-dsRNA ends, did not reduce the constitutive activation of ATP hydrolysis-deficient RIG-I (***Figure 1C***). However, RD and ATP binding are clearly important for signaling, as shown by △RD and K270I mutations by us and others (***Louber et al., 2015***) (***Figure 1C***), suggesting that a ring-like, ATP-bound structure is also involved in signaling caused by self-RNA (***Figure 6C***). In this conformation, the RD likely helps to displace the 2CARD module from the SF2 domain but may not have a high affinity for the RNA itself. Of note, the mutation in V699 of motif V also leads to increased constitutive signaling (***Figure 1C***). A plausible explanation could be that this mutation in RecA2 decouples RNA-binding induced ATP hydrolysis from translocation or displacement of RNA. In summary, our results suggest a model where RIG-I's translocase removes SF2 from dsRNA, perhaps at nearby bulges, unless high-affinity binding by the RD on RNA ends containing di- or triphosphates tethers RIG-I despite ATP-hydrolysis and leads to repeated or prolonged exposure of the 2CARD module.

An unexpected finding was that trapping the ATP state of RIG-I leads to a particularly increased interaction with the large ribosomal subunit via the expansion segment ES7L (***Figure 3D, F***). This expansion segment is present in metazoan ribosomes, however its length is substantially increased in human compared to drosophila ribosomes. The function of these expansion segments is not understood, but since helix E (ES7L-E) was recently found to interact with the selenoprotein synthesis factor SBP2, it is likely that the RNA in these elements is accessible to cytosolic proteins (***Kossinova et al., 2014***). The specific enrichment of the large ribosomal subunit under conditions where ribosomal subunits disengage argues for rather specific interactions of RIG-I E373Q with RNA present on the large but not the small subunit. The dominant binding of ribosomes by RIG-I E373Q can be explained by the high abundance of ribosomal RNA compared to other potential RIG-I ligands in the cytosol. We could directly visualize RIG-I E373Q on the ribosome at the solvent exposed root of ES7L-A (***Figure 3F, G***). This site contains a stretch of seven G:C/C:G base pairs, which approximately matches the footprint of dsRNA across the two SF2 RecA domains in the crystal structure of ADP·BeF_x-bound RIG-I (***Jiang et al., 2011***; ***Kohlway et al., 2013***; ***Kowalinski et al., 2011***; ***Luo et al., 2011***) and also meets the requirements for activation of RIG-I's ATPase

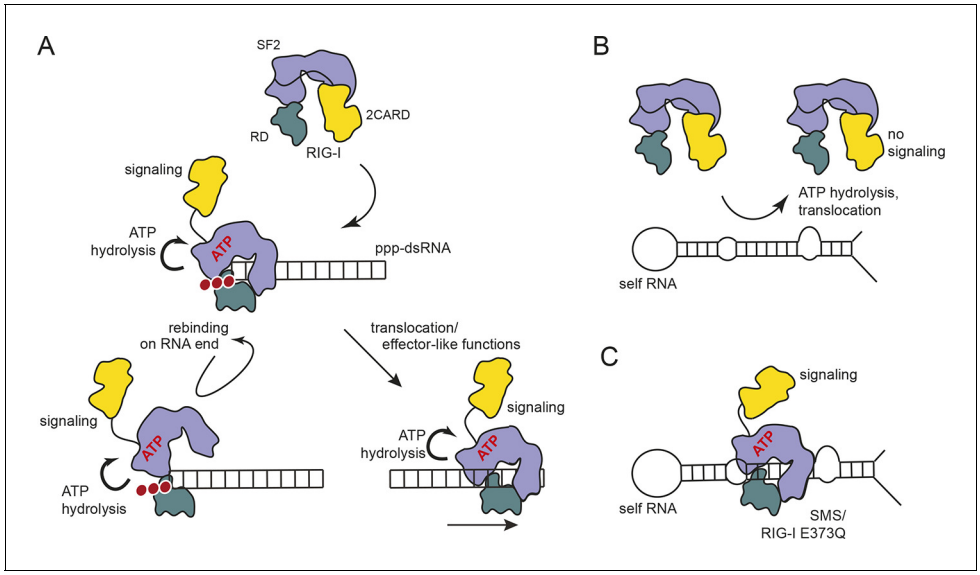


Figure 6. Proposed model for impact of ATP on RIG-I signaling on different RNAs. (**A**) RIG-I recognizes tri- or diphosphorylated double-stranded RNA and preferentially binds to the RNA end through its regulatory domain (RD, green). Binding of ATP-SF2 (purple) to the dsRNA releases the 2CARD module (yellow) and activates the downstream signaling process. ATP hydrolysis displaces the SF2 domain from dsRNA leading to either rebinding at the RNA end (tethered by RD) or to translocation along the RNA. (**B**) In healthy cells, sustained binding of RIG-I to self-RNA containing dsRNA stretches is prevented by ATP hydrolysis. The SF2 domain can be sufficiently displaced because the RD does not provide a high affinity tether. (**C**) Mutations that allow ATP promoted binding of dsRNA and displacement of the 2CARD module, but prevent ATP hydrolysis dependent dissociation of SF2 from dsRNA, such as those underlying atypical Singleton-Merten Syndrome, will result in an unintended signaling through self-RNA.

DOI: 10.7554/eLife.10859.015

(***Anchisi et al., 2015***). Since 40% of the particles had this additional density, it is conceivable that additional binding sites could contribute to the interaction with RIG-I E737Q as well. However, the peripheral parts of the expansion segments are flexible and not visible in the 3D reconstructions, preventing us from observing RIG-I at other regions.

The RNA corresponding to the observed binding region of ES7L-A is also bound by RIG-I in vitro and can moderately stimulate RIG-I's ATPase (***Figure 4D***, ***Figure 5A***). The much more efficient stimulation of RIG-I's ATPase by ppp-dsRNA is likely due to the high affinity towards RD, which could repeatedly "present" the RNA to SF2 (i.e. increasing the "local" concentration of RNA at SF2). Of note, while the addition of ATP to RIG-I reduces the interaction with the ES hairpin RNA, consistent with a role of the ATPase in preventing interaction with self-RNA, RIG-I E373Q binds with a moderately increased affinity to the ES-hairpin RNA in presence of ATP. Because of the large number of ribosomes in the cytosol it is therefore conceivable that RIG-I binds to double-stranded ribosomal RNA, including ES7L-A, under conditions where the ATPase is not able to efficiently displace the protein, such as those arising in patients with atypical SMS. In addition, the high local concentration of ribosomes in polysomes as well as a potential binding of RIG-I to other expansion segments could bring multiple RIG-I E373Q in contact, such that their exposed 2CARD module could interact for downstream signaling (***Peisley et al., 2014***; ***Wu et al., 2014***). We do not, however, want to rule out contributions by other self-ligands as well. For instance, RIG-I can bind to endogenous mRNA (***Zhang et al., 2013***) or RNase-L cleavage products (***Malathi et al., 2007***), while MDA5 was shown to be activated by mRNA stem loop structures under conditions where reduction of A:T base-paired RNA is not prevented by ADAR1 (***Liddicoat et al., 2015***).

In any case, there are two levels of control to limit RLR mediated signaling to viral RNA. On one hand, RNA editing (***Liddicoat et al., 2015***) and methylation (***Schuberth-Wagner et al., 2015***) modifies particular types of self-RNA that would otherwise form reasonable ligands for RIG-I or MDA5. On the other hand, the intrinsic ATPase and translocase activity removes RLRs from short, but abundant endogenous dsRNA stretches, thereby reducing background signaling and increasing the sensitivity of the system.

Materials and methods

Cell lines, viruses and antibodies

Luciferase assays and RIG-I:RNA co-immunopurifications were carried out in HEK 293T cells (purchased from ATCC, CRL-11268) or HEK 293T RIG-I KO cells (**Zhu et al., 2014**). HEK 293T ISRE-FF/RFP reporter cells (stable expression of firefly luciferase and RFP under the control of an ISRE promoter, kindly provided by Luis Martinez-Sorbid, University of Rochester, Rochester, NY) were used for interferon stimulated luciferase reporter gene assays of recovered RNA. HEK cells were maintained in high glucose Dulbecco's Modified Eagle Medium supplemented with GlutaMAX, pyruvate and 10% FBS (all purchased from Gibco, UK). Human ribosomes were purified from HeLa S3 cells cultured in SMEM (Sigma, Germany) supplemented with 10% FBS, Penicillin (100 U/mL)/ Streptamycin (100 µg/mL) and 1x GlutaMAX (all purchased from Gibco, UK) using a spinner flask at 40 rpm. All cell lines were routinely checked for Mycoplasms by PCR and were, except for the HEK 293T ISRE-FF/RFP cell line, tested to be negative. Mycoplasm contaminations were suppressed using Plasmocin (InvivoGen, France) according to the manufacturer's protocol. Viruses used for infections were Sendai virus Cantell, Sendai virus defective interfering particles H4 (kindly provided by Dominique Garcin, Geneva, Switzerland), recombinant measles virus (MeV) with a sequence identical to the vaccine strain Schwarz (AF266291.1.)(**del Valle et al., 2007**; **Devaux et al., 2007**) and recombinant MeV-Cko-ATU-Cs. MeV-Cko-ATU-Cs expresses the C Schwarz protein from an additional transcription unit (ATU) located between the M and the P gene, while expression of C from the P gene is abrogated. Specifically, three stop codons were introduced into the P gene for the C ORF while leaving P and V protein expression intact. Cloning was done as described previously (**Pfaller and Conzelmann, 2008**; **Sparrer et al., 2012**). Additionally, an ATU was introduced between the P and M gene by duplicating the gene borders of the P gene. The ORF of the C (Schwarz) protein was cloned into that ATU and the virus rescued from cDNA using helper plasmids in 293-3-46 cells (**Radecke et al., 1995**) and propagated on Vero cells as described previously (**Parks et al., 1999**; **Pfaller et al., 2014**). Primary antibodies to human MDA5 (AT113) and RIG-I (Alme-1) were purchased from Enzo Life Science (Loerrach, Germany). Antibodies to FLAG (M2), HA (HA-7) and β-tubulin (TUB 2.1) were obtained from Sigma-Aldrich (Saint Luis, MO, USA). Secondary antibodies were supplied by GE Healthcare (Buckinghamshire, UK).

Generation of RLR mutants

Sequences encoding full-length human RIG-I or MDA5 with N- or C-terminal FLAG/HA-tag were cloned into pcDNA5 FRT/TO (Invitrogen, Carlsbad, CA, USA). Mutants were generated by site-directed mutagenesis with PfuUltra polymerase (Agilent, Santa Clara, CA, USA).

Immunoprecipitation of RLR-associated RNA from infected or non-infected cells

$6x10^6$ HEK 293T or HEK 293T RIG-I KO cells were transfected with 10 µg pcDNA5 vector coding for different FLAG/HA tagged RLR proteins. Non-infected cells were harvested 24 h after transfection. Infections were carried out 6h after transfection with an MOI of 0.05 for measles virus or high MOI for Sendai virus and were allowed to proceed for 40 or 24 hr, respectively. Cells were harvested and incubated in Nonidet P-40 lysis buffer (50 mM HEPES, 150 mM KCl, 1 mM NaF, 0.5% NP-40, 0.5 mM DTT, protease inhibitor (Sigma, Saint Luis, MO, USA), pH 7.5) for 10 min on ice. Lysates were cleared by centrifugation and proteins were immunoprecipitated for 2.5 - –4 hr with anti-DDK magnetic beads (OriGene, Rockville, MD, USA) or anti-FLAG (M2) bound to magnetic protein G Dynabeads (Novex, Life Technologies, Carlsbad, CA, USA). Beads were washed five times with washing buffer (50 mM HEPES, 300 mM KCl, 0.05% NP-40, 0.5 mM DTT, protease inhibitor, pH 7.5) and incubated with proteinase K (Thermo Scientific, Vilnius, Lithuania) for 30 min at 50 °C. RNA was isolated by phenol/ chloroform/ isoamyl alcohol extraction using Phase Lock Gel Heavy tubes (5 PRIME, Germany). The quality of the isolated RNA was validated on an Agilent RNA 6000 Nano chip.

Luciferase transfection assays

Immunoactivity experiments were carried out in 24-well plates seeded with 2.5×10^5 HEK 293T RIG-I KO or 2.5×10^5 HEK 293T ISRE-FF/RFP reporter cells per well using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) as transfection reagent according to the manufacturer's protocol. For downstream signaling assays HEK 293T RIG-I KO cells were co-transfected with 500 ng protein expression vector, 100 ng p125-luc, 10 ng pCMV-RL and 50 ng empty expression vector. For RIG-I E373Q/RIG-I Δ2CARD,E373Q competition assays HEK 293T RIG-I KO cells were co-transfected with 100 ng RIG-I E373Q expression vector, varying concentrations of the RIG-I Δ2CARD,E373Q expression vector, 100 ng p125-luc and 10 ng pCMV-RL. DNA concentrations were held constant by adding empty expression vector if necessary. For determination of the immunostimulatory potential of recovered RNA from co-immunoprecipitations, HEK 293T ISRE-FF/RFP cells were transfected with 250 ng RNA in Opti-MEM (Gibco, UK). For RNA stimulation of cells overexpressing RIG-I E373Q 2.5×10^5 HEK 293T RIG-I KO cells were transfected with 100 ng RIG-I E373Q expression vector, 100 ng p125-luc, 10 ng pCMV-RL and 1000 ng total RNA/ rRNA or ES hairpin RNA in Opti-MEM. All cells were harvested 24 h after transfection using 200 µL PLB (Promega, Madison, WI, USA) and subjected to immunoactivity experiments using the Dual-Glo luciferase assay system (Promega, Madison, WI, USA) as previously described (***Runge et al., 2014***). The luciferase activity was determined with a Berthold Luminometer in 96-well plates using 20 µL cell lysate.

Protein expression and purification

RIG-I and RIG-I E373Q were expressed and purified from insect cells as described previously (***Cui et al., 2008***). Briefly, sequences encoding RIG-I were cloned into pFBDM vectors and transformed into *E. coli* DH10MultiBac cells. Bacmids were extracted for transfection into SF9 insect cells and propagated virus was used for protein expression in High Five insect cells. Seventy-two hours after infection cells were harvested and flash frozen in liquid nitrogen. RIG-I Δ2CARD was expressed in *E. coli* BL21 Rosetta (DE3), using pET expression vectors as described earlier (***Cui et al., 2008***). All recombinant proteins were purified using metal affinity (QIAGEN, Germany), heparin affinity and gel filtration chromatography (both GE Healthcare, Buckinghamshire, UK). Fractions containing RIG-I were concentrated to 6 mg/mL and flash-frozen in liquid nitrogen.

Thermal unfolding assay

Thermal stability of RIG-I or RIG-I E373Q in presence or absence of ATP was analyzed by fluorescence thermal shift assays. Proteins (20 µM) were incubated in 25 mM HEPES pH 7, 150 mM NaCl, 10 mM $MgCl_2$, 5 mM TCEP, 5% glycerol and 5 mM ATP. After addition of SYPRO orange (Invitrogen, Carlsbad, CA, USA, final concentration: 2.5x) the fluorescence signal was detected using a gradient from 5 °C to 100 °C with 0.5 K/30 s and one scan each 0.5 K in a real-time thermal cycler (Biorad, Germany, CFX96 touch) using the FRET mode.

Small-angle X-ray scattering

SAXS experiments were conducted at the PETRA3 P12 beamline of the European Molecular Biology Laboratory/ Deutsches Elektronen-Synchrotron, Hamburg, Germany. Samples were measured in absence or presence of 5 mM ATP in size exclusion buffer (25 mM HEPES pH 7, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM β-Mercaptoethanol, 5% glycerol). RIG-I samples were measured at protein concentrations of 1.28, 2.65 and 8.35 mg/mL and RIG-I E373Q samples with concentrations of 0.87, 2.13 and 6.84 mg/mL. The respective scattering of the corresponding buffer was used for buffer subtraction. The samples did not show signs of radiation damage, which was assessed by automatic and manual comparison of consecutive exposure frames. The data was processed using PRIMUS from the ATSAS package (***Konarev et al., 2006***) and the radius of gyration was determined by Guinier plot [ln I(s) versus s^2] analysis obeying the Guinier approximation for globular proteins ($s \times R_g < 1.3$).

Human 80S ribosome preparation

HeLa S3 cells were harvested (2 min, 650 x g), washed with PBS (Invitrogen, Carlsbad, CA, USA) and incubated with 1.5x vol Buffer 1 (10 mM HEPES/KOH, pH 7.2/4 °C, 10 mM KOAc, 1 mM $Mg(OAc)_2$ and 1 mM DTT) for 15 min on ice, followed by disruption with nitrogen pressure (300 psi, 30 min, 4 °C) in a cell disruption vessel (Parr Instrument, Moline, IL, USA). The cell lysate was cleared (10 min,

14,000 rpm, Eppendorf 5417R, 4 °C) and the resulting supernatant was loaded onto a sucrose cushion (Buffer 1 supplemented with 35% sucrose). Subsequent spinning (98 min, 75.000 rpm, TLA 120.2, 4 °C) was performed. After resuspension of the ribosomal pellet, a high-salt purification by centrifugation through a 500 mM sucrose cushion (50 mM Tris/HCl, pH 7.0/4 °C, 500 mM KOAc, 25 mM $Mg(OAc)_2$, 5 mM β-mercaptoethanol, 1 M sucrose, 1 µg/mL cycloheximide and 0.1% Nikkol) was conducted (45 min, 100,000 rpm, TLA120.2, 4 °C). The ribosomal pellet was resuspended in Ribosome Buffer (50 mM Tris/HCl, pH 7.0/4 °C, 100 mM KOAc, 6 mM $Mg(OAc)_2$, 1 mM DTT, 1/200 EDTA-free Complete protease inhibitor (Roche, Germany), 0.2 U/mL RNasin (Promega, Madison, WI, USA)), quickly centrifuged, frozen in liquid nitrogen and stored at -80 °C.

Total RNA and ribosomal RNA isolation

For total RNA isolation 2.5 x 10^5 HEK 293T were seeded per well of 24 well plates. After 24 h cells were harvested in PBS, collected by centrifugation and lysed in Nonidet P-40 lysis buffer for 10 min on ice. Supernatant was cleared by centrifugation and DNA was digested with TURBO DNase (Ambion, Life Technologies, Carlsbad, CA, USA) for 3 min at 37 °C. Proteins were digested and RNA was extracted as described above. For ribosomal RNA isolation purified human ribosomes were proteinase K digested and RNA was extracted accordingly.

Ribosomal binding studies

Human 80S ribosomes were incubated with or without 2.5x molar excess of RIG-I or RIG-I E373Q in binding buffer (50 mM HEPES/KOH, pH 7.5/ 4 °C, 100 mM KCl, 2.5 mM $Mg(OAc)_2$, 2 mM DTT, 1 mM ATP, 0.1% DDM, 10% Glycerol) for 15 min at room temperature and then for 15 min at 4 °C. The mixture was loaded onto a sucrose cushion (binding buffer with 750 mM sucrose) and spun (3 h, 40,000 rpm, SW55Ti, 4 °C). Supernatant and pellet fractions were separated and TCA precipitated. The resulting samples were analyzed by SDS-PAGE and visualized using SYPRO Orange Staining (Molecular Probes, Eugene, OR, USA).

Cryo-grid preparation

5 OD/mL human 80S ribosomes were incubated with or without 2.5x molar excess of RIG-I E373Q. Each sample (50 mM HEPES / KOH, pH 7.5 / 4 °C, 100 mM KCl, 2.5 mM $Mg(OAc)_2$, 2 mM DTT, 1 mM ATP, 0.1% DDM, 5% glycerol) was applied to 2 nm pre-coated Quantifoil R3/3 holey carbon supported grids and vitrified using a Vitrobot Mark IV (FEI Company , Germany).

Cryo-electron microscopy and single particle reconstruction

Data were collected on a 120 keV TECNAI SPIRIT cryo-electron microscope with a pixel size of 2.85 Å/pixel at a defocus range between 1.4 µm and 4.6 µm (with RIG-I E373Q ligand) or between 1.8 µm and 5.3 µm (without ligand) under low dose conditions. Particles were detected with SIGNATURE (***Chen and Grigorieff, 2007***). Initial alignment resulted in 61,067 particles (with ligand) and 29,959 particles (without ligand). Subsequent data processing and single particle analysis was performed using the SPIDER software package (***Frank et al., 1996***). Non-ribosomal particles (19,080 particles, 31% (with ligand) and 10,663 particles, 35% (without ligand)) were removed from each data set by unsupervised 3D sorting (***Loerke et al., 2010***). The remaining particles were further sorted, resulting in a volume with additional density (with ligand: 23,715 particles, 39%). The identical sorting scheme was applied to the control 80S ribosome without ligand, resulting in final 11,727 particles (39%). The final 80S structures with and without ligand were refined to an overall resolution ($FCS_{0.5}$) of 17.7 Å and 21.9 Å, respectively. For comparison of the two final volumes, a statistical difference map between the two reconstructions was calculated.

Figure preparations and model docking

We used the crystal structure of the human RIG-I protein (PDB code 3TMI) (***Jiang et al., 2011***) and the human ribosome (PDB 4V6X) (***Anger, et al., 2013***) for rigid-body fitting into the additional density. Figures depicting atomic models with and without density were prepared using UCSF Chimera (***Pettersen et al., 2004***).

Differential radial capillary action of ligand assay

ATP binding was determined by DRaCALA using [α-^{32}P]ATP (Hartmann Analytik, Germany). 12 μM RIG-I or RIG-I E373Q were incubated in 50 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl2, 2.5 mM TCEP, 0.1 mg/mL BSA supplemented with 2.5 nM [α-^{32}P]ATP for 10 min at room temperature in presence or absence of 100 nM RNA. 2.5 μL of reaction mixture was spotted on nitrocellulose membranes (0.22 μM pores, GE Healthcare, Buckinghamshire, UK), air-dried and [α-^{32}P]ATP was detected using a phosphor-imaging system (GE Healthcare, Germany).

Electrophoretic mobility shift assay

Proteins at different concentrations were pre-incubated with ATP, ADP or ADP·BeF_3 (all 3 mM end concentration, ADP·BeF_3 was generated using ADP, NaF and $BeCl_2$ in a 1:1:5 molar ratio) and added to 0.5 μM ES hairpin RNA or 0.2 μM 24mer RNA in EMSA buffer (50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 5 mM TCEP, 7.5 μM $ZnCl_2$, 3 mM ATP, 5% glycerol). Reactions were incubated for 20 min at 37 °C. Samples were separated on TB agarose gels (89 mM Tris, 89 mM boric acid, 0.8% agarose) and stained with Gel-Red (Biotium, Hayward, CA, USA). Unbound RNA bands were quantified with ImageJ.

Fluorescence anisotropy assays

Different RIG-I or RIG-I E373Q protein concentrations were titrated into EMSA buffer without ATP and glycerol. Reactions were started by addition of 5 mM ATP and 20 nM Cy3- or Cy5-labeled RNA and fluorescence anisotropy was measured with a TECAN M1000 plate reader after incubation at room temperature for 20 min.

ATPase hydrolysis assays

ATPase hydrolysis activity was determined using [γ-^{32}P]ATP (Hartmann Analytik, Germany). Proteins at different concentrations were pre-incubated with 100 nM RNA or purified ribosomes for 10 min at room temperature in EMSA buffer without ATP. The reaction was initiated by addition of 1.5 mM unlabeled and 10 nM [γ-^{32}P]ATP and incubated for 20 min at 37 °C. Free phosphate was separated from ATP by thin layer chromatography in TLC running buffer (1 M formic acid, 0.5 M LiCl) on polyethyleneimine cellulose TLC plates (Sigma-Aldrich, Germany). [γ-^{32}P]P_i and [γ-^{32}P]ATP were detected using a phosphor-imaging system (GE Healthcare, Germany) and quantified using ImageJ.

Acknowledgements

We thank Simon Runge for help with RNA co-purification during initial stages of the project and Filiz Civril for the RIG-I insect cell expression vector. We thank Luis Martinez-Sorbido (University of Rochester, Rochester, NY) for the HEK 293T ISRE-FF/RFP cell line and Dominique Garcin (University of Geneva, Switzerland) for providing SeV defective interfering particles H4. We thank Stefan Krebs and Andrea Klanner for support in RNA analysis and Andrea Gilmozzi for help with ribosome purification and binding assays. Furthermore, we thank the staff of the EMBL/DESY PETRA3 P12 beamline for support in SAXS measurements, Gregor Witte for help with SAXS data analysis, Tobias Deimling for help with EMSAs and ATPase assays and Robert Byrne for critically reading the manuscript.

Additional information

Competing interests

GH, VH co-founder and shareholder of the Rigontec GmbH. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
National Institutes of Health	T32 training grant 5T32AI007647-13	Jenish R Patel
German Excellence Initiative	CIPSM	Roland Beckmann Karl-Peter Hopfner

Graduate School of Quantitative Biosciences Munich		Roland Beckmann Karl-Peter Hopfner
European Research Council	Advanced Grant CRYOTRANSLATION	Roland Beckmann
Deutsche Forschungsgemeinschaft	FOR 1805	Roland Beckmann
Deutsche Forschungsgemeinschaft	SFB646	Roland Beckmann Karl-Peter Hopfner
Deutsche Forschungsgemeinschaft	GRK1721	Karl-Peter Hopfner Roland Beckmann
Bavarian network for Molecular Biosystems		Karl-Peter Hopfner

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

CL, Planned and performed luciferase assays and co-IP experiments, purified proteins, performed in vitro assays and wrote the paper; SM, Purified human ribosomes, carried out ribosome binding studies and analyzed the cryo-EM structure; KMJS, Produced measles virus and planned the luciferase assays and co-IPs; CCdOM, Carried out co-IP experiments and thermal shift assays; MM, Purified proteins; JRP, Provided Sendai virus Cantell; MG, GH, Provided synthetic 24mer and 12mer ppp-RNAs; AGS, K-KC, RB, Supervised the study and critically read the manuscript; VH, Provided the 293T RIG-I KO cell line and critically read the manuscript; K-PH, Designed the research, supervised experiments and wrote the paper

Author ORCIDs

Karl-Peter Hopfner, http://orcid.org/0000-0002-4528-8357

Additional files

Major datasets

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Jiang F, Ramanathan A, Miller MT, Tang GQ, Gale M, Patel SS, Marcotrigiano J	2011	Structural basis of RNA recognition and activation by innate immune receptor RIG-I	http://www.rcsb.org/pdb/explore/explore.do?structureId=3TMI	Publicly available at the RCSB Protein Data Bank (Accession no: 3TMI).
Anger AM, Armache JP, Beringhausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R	2013	Structures of the human and Drosophila 80S ribosome	http://www.rcsb.org/pdb/explore/explore.do?structureId=4V6X	Publicly available at the RCSB Protein Data Bank (Accession no: 4V6X).

References

Anchisi S, Guerra J, Garcin D. 2015. RIG-I ATPase activity and discrimination of self-RNA versus non-self-RNA. *mBio* **6**:e02349-14. doi: 10.1128/mBio.02349-14

Anger AM, Armache J-P, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. 2013. Structures of the human and drosophila 80S ribosome. *Nature* **497**:80–85. doi: 10.1038/nature12104

Bamming D, Horvath CM. 2009. Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *Journal of Biological Chemistry* **284**:9700–9712. doi: 10.1074/jbc.M807365200

Banerjee AK, Shatkin AJ. 1971. Guanosine-5′-diphosphate at the 5′ termini of reovirus RNA: evidence for a segmented genome within the virion. *Journal of Molecular Biology* **61**:643–653. doi: 10.1016/0022-2836(71)90069-6

Berke IC, Modis Y. 2012. MDA5 cooperatively forms dimers and ATP-sensitive filaments upon binding double-stranded RNA. *The EMBO Journal* **31**:1714–1726. doi: 10.1038/emboj.2012.19

Brubaker SW, Bonham KS, Zanoni I, Kagan JC. 2015. Innate immune pattern recognition: a cell biological perspective. *Annual Review of Immunology* **33**:257–290. doi: 10.1146/annurev-immunol-032414-112240

Bruns AM, Leser GP, Lamb RA, Horvath CM. 2014. The innate immune sensor LGP2 activates antiviral signaling by regulating MDA5-RNA interaction and filament assembly. *Molecular Cell* **55**:771–781. doi: 10.1016/j.molcel.2014.07.003

Cedergreen N, Ritz C, Streibig JC. 2005. Improved empirical models describing hormesis. *Environmental Toxicology and Chemistry* **24**:3166–3172. doi: 10.1897/05-014R.1

Chen JZ, Grigorieff N. 2007. SIGNATURE: a single-particle selection system for molecular electron microscopy. *Journal of Structural Biology* **157**:168–173. doi: 10.1016/j.jsb.2006.06.001

Civril F, Bennett M, Moldt M, Deimling T, Witte G, Schiesser S, Carell T, Hopfner K-P. 2011. The RIG-I ATPase domain structure reveals insights into ATP-dependent antiviral signalling. *EMBO Reports* **12**:1127–1134. doi: 10.1038/embor.2011.190

Cui S, Eisenächer K, Kirchhofer A, Brzózka K, Lammens A, Lammens K, Fujita T, Conzelmann K-K, Krug A, Hopfner K-P. 2008. The C-terminal regulatory domain is the RNA 5′-triphosphate sensor of RIG-I. *Molecular Cell* **29**:169–179. doi: 10.1016/j.molcel.2007.10.032

del Valle JR, Devaux P, Hodge G, Wegner NJ, McChesney MB, Cattaneo R. 2007. A vectored measles virus induces hepatitis B surface antigen antibodies while protecting macaques against measles virus challenge. *Journal of Virology* **81**:10597–10605. doi: 10.1128/JVI.00923-07

Devaux P, von Messling V, Songsungthong W, Springfeld C, Cattaneo R. 2007. Tyrosine 110 in the measles virus phosphoprotein is required to block STAT1 phosphorylation. *Virology* **360**:72–83. doi: 10.1016/j.virol.2006.09.049

Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *Journal of Structural Biology* **116**:190–199. doi: 10.1006/jsbi.1996.0030

Funabiki M, Kato H, Miyachi Y, Toki H, Motegi H, Inoue M, Minowa O, Yoshida A, Deguchi K, Sato H, Ito S, Shiroishi T, Takeyasu K, Noda T, Fujita T. 2014. Autoimmune disorders associated with gain of function of the intracellular sensor MDA5. *Immunity* **40**:199–212. doi: 10.1016/j.immuni.2013.12.014

Gack MU, Shin YC, Joo C-H, Urano T, Liang C, Sun L, Takeuchi O, Akira S, Chen Z, Inoue S, Jung JU. 2007. TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446**:916–920. doi: 10.1038/nature05732

Goldeck M, Tuschl T, Hartmann G, Ludwig J. 2014. Efficient solid-phase synthesis of pppRNA by using product-specific labeling. *Angewandte Chemie* **126**:4782–4786. doi: 10.1002/ange.201400672

Goubau D, Schlee M, Deddouche S, Pruijssers AJ, Zillinger T, Goldeck M, Schuberth C, Van der Veen AG, Fujimura T, Rehwinkel J, Iskarpatyoti JA, Barchet W, Ludwig J, Dermody TS, Hartmann G, Reis e Sousa C. 2014. Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5′-diphosphates. *Nature* **514**:372–375. doi: 10.1038/nature13590

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The vienna RNA websuite. *Nucleic Acids Research* **36**:W70–W74. doi: 10.1093/nar/gkn188

Hornung V, Ellegast J, Kim S, Brzozka K, Jung A, Kato H, Poeck H, Akira S, Conzelmann K-K, Schlee M, Endres S, Hartmann G. 2006. 5′-triphosphate RNA is the ligand for RIG-I. *Science* **314**:994–997. doi: 10.1126/science.1132505

Hou F, Sun L, Zheng H, Skaug B, Jiang Q-X, Chen ZJ. 2011. MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* **146**:448–461. doi: 10.1016/j.cell.2011.06.041

Iwasaki A, Medzhitov R. 2015. Control of adaptive immunity by the innate immune system. *Nature Immunology* **16**:343–353. doi: 10.1038/ni.3123

Jang M-A, Kim EK, Now H, Nguyen NTH, Kim W-J, Yoo J-Y, Lee J, Jeong Y-M, Kim C-H, Kim O-H, Sohn S, Nam S-H, Hong Y, Lee YS, Chang S-A, Jang SY, Kim J-W, Lee M-S, Lim SY, Sung K-S, Park K-T, Kim BJ, Lee J-H, Kim D-K, Kee C, Ki C-S. 2015. Mutations in DDX58, which encodes RIG-I, cause atypical Singleton-Merten syndrome. *The American Journal of Human Genetics* **96**:266–274. doi: 10.1016/j.ajhg.2014.11.019

Jiang F, Ramanathan A, Miller MT, Tang G-Q, Gale M, Patel SS, Marcotrigiano J. 2011. Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* **479**:423–427. doi: 10.1038/nature10537

Kato H, Takeuchi O, Mikamo-Satoh E, Hirai R, Kawai T, Matsushita K, Hiiragi A, Dermody TS, Fujita T, Akira S. 2008. Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene I and melanoma differentiation-associated gene 5. *Journal of Experimental Medicine* **205**:1601–1610. doi: 10.1084/jem.20080091

Kohlway A, Luo D, Rawling DC, Ding SC, Pyle AM. 2013. Defining the functional determinants for RNA surveillance by RIG-I. *EMBO Reports* **14**:772–779. doi: 10.1038/embor.2013.108

Konarev PV, Petoukhov MV, Volkov VV, Svergun DI. 2006. *ATSAS* 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **39**:277–286. doi: 10.1107/S0021889806004699

Kossinova O, Malygin A, Krol A, Karpova G. 2014. The SBP2 protein central to selenoprotein synthesis contacts the human ribosome at expansion segment 7L of the 28S rRNA. *RNA* **20**:1046–1056. doi: 10.1261/rna.044917.114

Kowalinski E, Lunardi T, McCarthy AA, Louber J, Brunel J, Grigorov B, Gerlier D, Cusack S. 2011. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**:423–435. doi: 10.1016/j.cell.2011.09.039

Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, Li JB, Seeburg PH, Walkley CR. 2015. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**:1115–1120. doi: 10.1126/science.aac7049

Liu G, Park H-S, Pyo H-M, Liu Q, Zhou Y, Lyles DS. 2015. Influenza a virus panhandle structure is directly involved in RIG-I activation and interferon induction. *Journal of Virology* **89**:6067–6079. doi: 10.1128/JVI.00232-15

Loerke J, Giesebrecht J, Spahn CMT. 2010. Multiparticle cryo-EM of ribosomes. *Methods Enzymol* **483**:161–177. doi: 10.1016/S0076-6879(10)83008-3

Louber J, Brunel J, Uchikawa E, Cusack S, Gerlier D. 2015. Kinetic discrimination of self/non-self RNA by the ATPase activity of RIG-I and MDA5. *BMC Biology* **13**. doi: 10.1186/s12915-015-0166-9

Luo D, Ding SC, Vela A, Kohlway A, Lindenbach BD, Pyle AM. 2011. Structural insights into RNA recognition by RIG-I. *Cell* **147**:409–422. doi: 10.1016/j.cell.2011.09.023

Malathi K, Dong B, Gale M, Silverman RH. 2007. Small self-RNA generated by RNase l amplifies antiviral innate immunity. *Nature* **448**:816–819. doi: 10.1038/nature06042

Myong S, Cui S, Cornish PV, Kirchhofer A, Gack MU, Jung JU, Hopfner K-P, Ha T. 2009. Cytosolic viral sensor RIG-I is a 5′-triphosphate-dependent translocase on double-stranded RNA. *Science* **323**:1070–1074. doi: 10.1126/science.1168352

Pandey S, Kawai T, Akira S. 2015. Microbial sensing by toll-like receptors and intracellular nucleic acid sensors. *Cold Spring Harbor Perspectives in Biology* **7**:a016246. doi: 10.1101/cshperspect.a016246

Parks CL, Lerch RA, Walpita P, Sidhu MS, Udem SA. 1999. Enhanced measles virus cDNA rescue and gene expression after heat shock. *Journal of Virology* **73**:3560–3566.

Patel JR, Jain A, Chou Yi-ying, Baum A, Ha T, García-Sastre A. 2013. ATPase-driven oligomerization of RIG-I on RNA allows optimal activation of type-I interferon. *EMBO Reports* **14**:780–787. doi: 10.1038/embor.2013.102

Peisley A, Wu B, Xu H, Chen ZJ, Hur S. 2014. Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I. *Nature* **509**:110–114. doi: 10.1038/nature13140

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**:1605–1612. doi: 10.1002/jcc.20084

Pfaller CK, Conzelmann K-K. 2008. Measles virus V protein is a decoy substrate for IkappaB kinase and prevents toll-like receptor 7/9-mediated interferon induction. *Journal of Virology* **82**:12365–12373. doi: 10.1128/JVI.01321-08

Pfaller CK, Radeke MJ, Cattaneo R, Samuel CE. 2014. Measles virus C protein impairs production of defective copyback double-stranded viral RNA and activation of protein kinase R. *Journal of Virology* **88**:456–468. doi: 10.1128/JVI.02572-13

Pichlmair A, Schulz O, Tan C-P, Rehwinkel J, Kato H, Takeuchi O, Akira S, Way M, Schiavo G, Reis e Sousa C. 2009. Activation of MDA5 requires higher-order RNA structures generated during virus infection. *Journal of Virology* **83**:10761–10769. doi: 10.1128/JVI.00770-09

Pichlmair A, Schulz O, Tan CP, Naslund TI, Liljestrom P, Weber F, Reis e Sousa C. 2006. RIG-I-mediated antiviral responses to single-stranded RNA bearing 5′-phosphates. *Science* **314**:997–1001. doi: 10.1126/science.1132998

Radecke F, Spielhofer P, Schneider H, Kaelin K, Huber M, Dötsch C, Christiansen G, Billeter MA. 1995. Rescue of measles viruses from cloned DNA. *The EMBO Journal* **14**.

Rice GI, del Toro Duany Y, Jenkinson EM, Forte GMA, Anderson BH, Ariaudo G, Bader-Meunier B, Baildam EM, Battini R, Beresford MW, Casarano M, Chouchane M, Cimaz R, Collins AE, Cordeiro NJV, Dale RC, Davidson JE, De Waele L, Desguerre I, Faivre L, Fazzi E, Isidor B, Lagae L, Latchman AR, Lebon P, Li C, Livingston JH, Lourenço CM, Mancardi MM, Masurel-Paulet A, McInnes IB, Menezes MP, Mignot C, O'Sullivan J, Orcesi S, Picco PP, Riva E, Robinson RA, Rodriguez D, Salvatici E, Scott C, Szybowska M, Tolmie JL, Vanderver A, Vanhulle C, Vieira JP, Webb K, Whitney RN, Williams SG, Wolfe LA, Zuberi SM, Hur S, Crow YJ. 2014. Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nature Genetics* **46**:503–509. doi: 10.1038/ng.2933

Rozen F, Pelletier J, Trachsel H, Sonenberg N. 1989. A lysine substitution in the ATP-binding site of eucaryotic initiation factor 4A abrogates nucleotide-binding activity. *Molecular and Cellular Biology* **9**:4061–4063. doi: 10.1128/MCB.9.9.4061

Runge S, Sparrer KMJ, Lässig C, Hembach K, Baum A, García-Sastre A, Söding J, Conzelmann K-K, Hopfner K-P, Mossman KL. 2014. In vivo ligands of MDA5 and RIG-I in measles virus-infected cells. *PLoS Pathogens* **10**:e1004081. doi: 10.1371/journal.ppat.1004081

Rutsch F, MacDougall M, Lu C, Buers I, Mamaeva O, Nitschke Y, Rice GI, Erlandsen H, Kehl HG, Thiele H, Nürnberg P, Höhne W, Crow YJ, Feigenbaum A, Hennekam RC. 2015. A specific IFIH1 gain-of-function mutation causes Singleton-Merten syndrome. *The American Journal of Human Genetics* **96**:275–282. doi: 10.1016/j.ajhg.2014.12.014

Satoh T, Kato H, Kumagai Y, Yoneyama M, Sato S, Matsushita K, Tsujimura T, Fujita T, Akira S, Takeuchi O. 2010. LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proceedings of the National Academy of Sciences of the United States of America* **107**:1512–1517. doi: 10.1073/pnas.0912986107

Schlee M, Roth A, Hornung V, Hagmann CA, Wimmenauer V, Barchet W, Coch C, Janke M, Mihailovic A, Wardle G, Juranek S, Kato H, Kawai T, Poeck H, Fitzgerald KA, Takeuchi O, Akira S, Tuschl T, Latz E, Ludwig J, Hartmann G. 2009. Recognition of 5′ triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunity* **31**:25–34. doi: 10.1016/j.immuni.2009.05.008

Schmidt A, Schwerd T, Hamm W, Hellmuth JC, Cui S, Wenzel M, Hoffmann FS, Michallet M-C, Besch R, Hopfner K-P, Endres S, Rothenfusser S. 2009. 5′-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 12067–12072. doi: 10.1073/pnas.0900971106

Schnell G, Loo Y-M, Marcotrigiano J, Gale M, Mossman KL. 2012. Uridine composition of the poly-U/UC tract of HCV RNA defines non-self recognition by RIG-I. *PLoS Pathogens* **8**:e1002839. doi: 10.1371/journal.ppat.1002839

Schuberth-Wagner C, Ludwig J, Bruder AK, Herzner A-M, Zillinger T, Goldeck M, Schmidt T, Schmid-Burgk JL, Kerber R, Wolter S, Stümpel J-P, Roth A, Bartok E, Drosten C, Coch C, Hornung V, Barchet W, Kümmerer BM, Hartmann G, Schlee M. 2015. A conserved histidine in the RNA sensor RIG-I controls immune tolerance to N1-2′O-methylated self RNA. *Immunity* **43**:41–51. doi: 10.1016/j.immuni.2015.06.015

Sparrer KMJ, Gack MU. 2015. Intracellular detection of viral nucleic acids. *Current Opinion in Microbiology* **26**: 1–9. doi: 10.1016/j.mib.2015.03.001

Sparrer KMJ, Pfaller CK, Conzelmann K-K. 2012. Measles virus C protein interferes with beta interferon transcription in the nucleus. *Journal of Virology* **86**:796–805. doi: 10.1128/JVI.05899-11

Wang Y, Ludwig J, Schuberth C, Goldeck M, Schlee M, Li H, Juranek S, Sheng G, Micura R, Tuschl T, Hartmann G, Patel DJ. 2010. Structural and functional insights into 5′-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nature Structural & Molecular Biology* **17**:781–787. doi: 10.1038/nsmb.1863

Weber M, Gawanbacht A, Habjan M, Rang A, Borner C, Schmidt AM, Veitinger S, Jacob R, Devignot S, Kochs G, García-Sastre A, Weber F. 2013. Incoming RNA virus nucleocapsids containing a 5′-triphosphorylated genome activate RIG-I and antiviral signaling. *Cell Host & Microbe* **13**:336–346. doi: 10.1016/j.chom.2013.01.012

Wu B, Peisley A, Tetrault D, Li Z, Egelman EH, Magor KE, Walz T, Penczek PA, Hur S. 2014. Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Molecular Cell* **55**:511–523. doi: 10.1016/j.molcel.2014.06.010

Wu J, Chen ZJ. 2014. Innate immune sensing and signaling of cytosolic nucleic acids. *Annual Review of Immunology* **32**:461–488. doi: 10.1146/annurev-immunol-032713-120156

Xu H, He X, Zheng H, Huang LJ, Hou F, Yu Z, de la Cruz MJ, Borkowski B, Zhang X, Chen ZJ, Jiang Q-X. 2014. Structural basis for the prion-like MAVS filaments in antiviral innate immunity. *eLife* **3**:e01489. doi: 10.7554/eLife.01489

Yao H, Dittmann M, Peisley A, Hoffmann H-H, Gilmore RH, Schmidt T, Schmid-Burgk JL, Hornung V, Rice CM, Hur S. 2015. ATP-dependent effector-like functions of RIG-I-like receptors. *Molecular Cell* **58**:541–548. doi: 10.1016/j.molcel.2015.03.014

Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, Taira K, Akira S, Fujita T. 2004. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunology* **5**:730–737. doi: 10.1038/ni1087

Zhang H-X, Liu Z-X, Sun Y-P, Zhu J, Lu S-Y, Liu X-S, Huang Q-H, Xie Y-Y, Zhu H-B, Dang S-Y, Chen H-F, Zheng G-Y, Li Y-X, Kuang Y, Fei J, Chen S-J, Chen Z, Wang Z-G, Lu SY, Liu XS, Huang QH, Xie YY, Zhu HB, Dang SY. 2013. Rig-I regulates NF- kappaB activity through binding to nf-kappab1 3′-UTR mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **110**:6459–6464. doi: 10.1073/pnas.1304432110

Zhu J, Zhang Y, Ghosh A, Cuevas RA, Forero A, Dhar J, Ibsen MS, Schmid-Burgk JL, Schmidt T, Ganapathiraju MK, Fujita T, Hartmann R, Barik S, Hornung V, Coyne CB, Sarkar SN. 2014. Antiviral activity of human OASL protein is mediated by enhancing signaling of the RIG-I RNA sensor. *Immunity* **40**:936–948. doi: 10.1016/j.immuni.2014.05.007

Figures and figure supplements

ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA

Charlotte Lässig *et al*

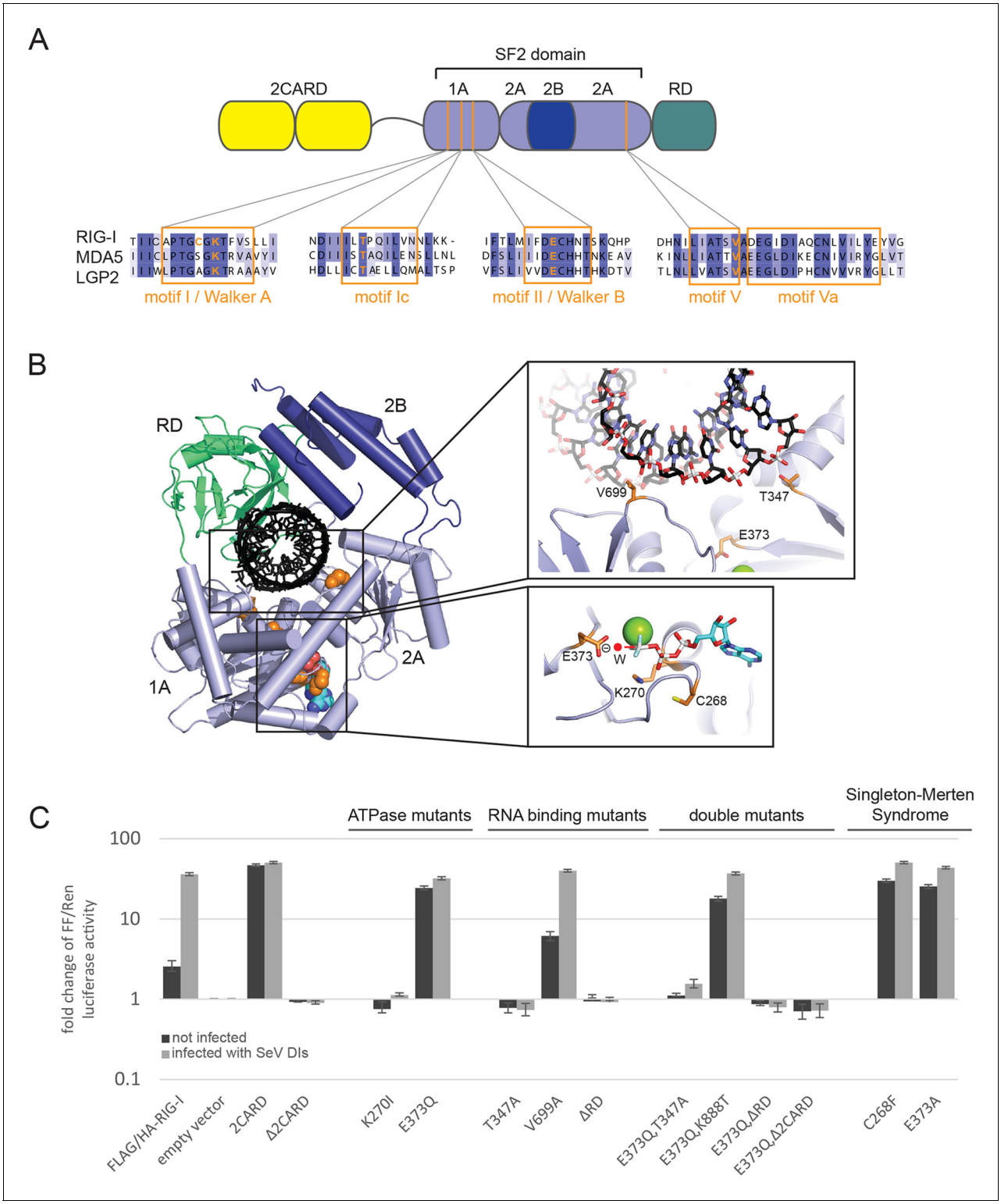


Figure 1. Cellular studies of RIG-I ATPase mutants in infected or non-infected cells. (**A**) Location of amino acid substitutions of RIG-I SF2 domain variants used in this study (orange lines) within different RLR helicase motifs (orange squares). (**B**) Single amino acid substitutions (orange) within the RIG-I 3D structure (PDB: 3TMI). (**C**) Fold change of interferon-β (IFNβ) promoter driven luciferase activity in uninfected HEK 293T RIG-I KO cells or in cells challenged with Sendai virus defective interfering particles (SeV DIs). Cells were co-transfected with RIG-I expression vectors and p125-luc/ pCMV-RL reporter plasmids, and infected with SeV DIs 6 hr post transfection. Firefly (FF) luciferase activities were determined in respect to Renilla (Ren) luciferase activities 24 hpi. All ratios were normalized to the empty vector control. n=3–12, error bars represent mean values ± standard deviation.
DOI: 10.7554/eLife.10859.003

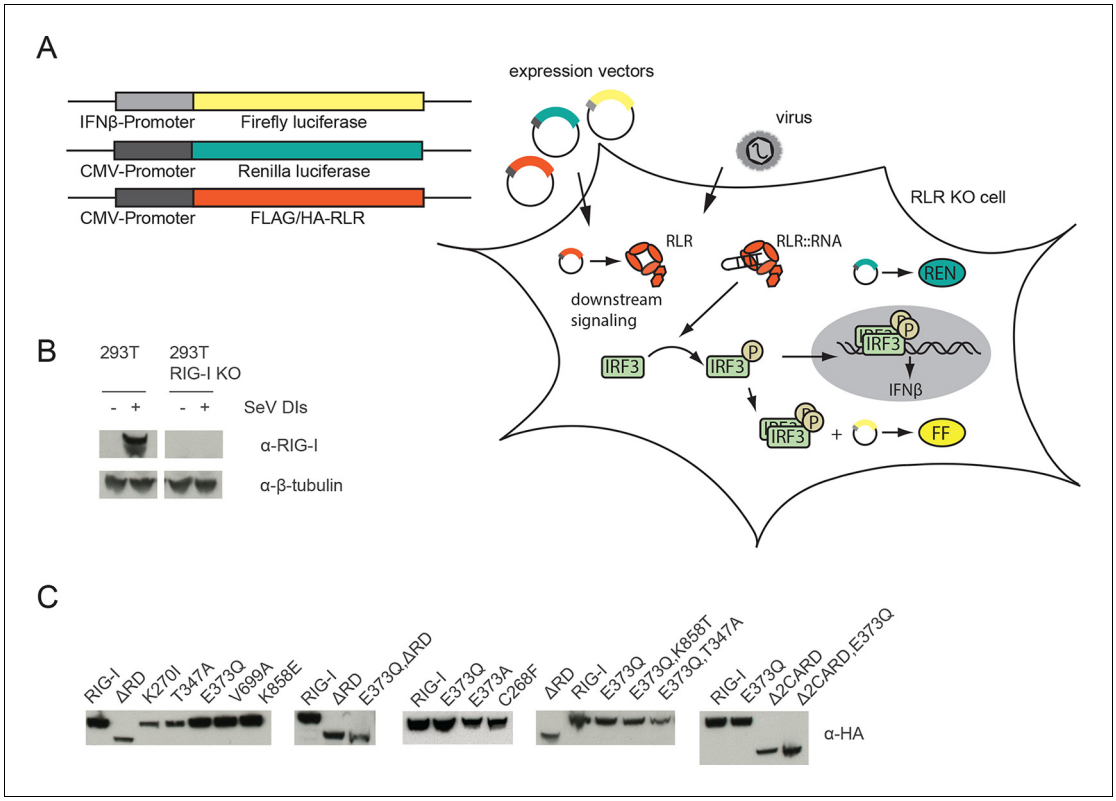


Figure 1—figure supplement 1. Assay for defining of the impact of RLR variant expression on RLR signaling in infected or non-infected cells. (**A**) HEK 293T RIG-I KO cells were co-transfected with different expression and control vectors as indicated. RLR signaling induces an interferon-β (IFNβ) promoter driven expression of firefly luciferase (FF). Renilla luciferase (Ren) is constitutively expressed via a CMV promoter and serves as transfection control. (**B**) Western Blot analysis of virus-induced RIG-I expression in HEK 293T and HEK 293T RIG-I KO cells. (**C**) Western Blot control of overexpressed FLAG/HA tagged RIG-I variants in HEK 293T RIG-I KO cells from ***Figure 1***, panel C.

DOI: 10.7554/eLife.10859.004

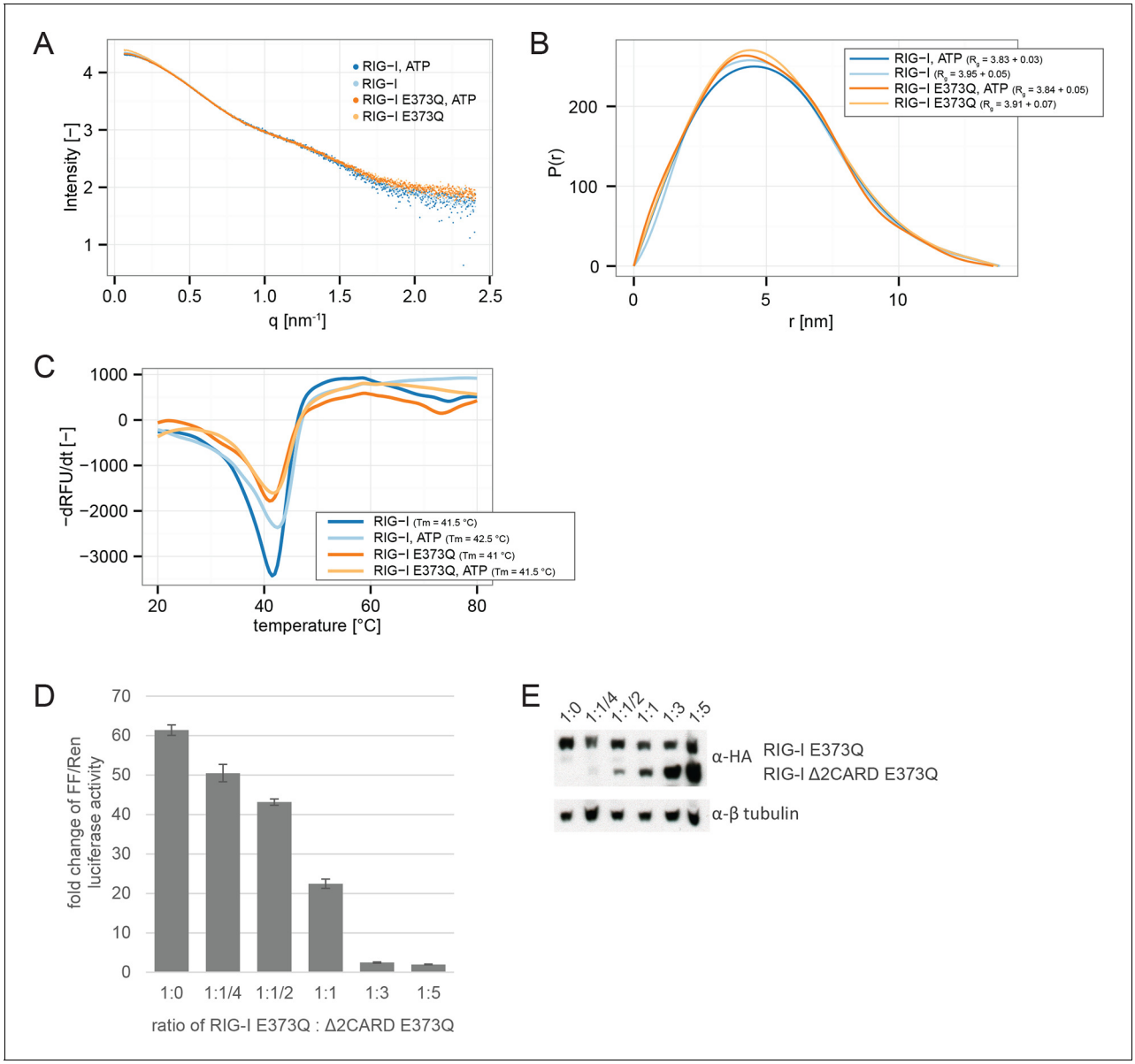


Figure 1—figure supplement 2. RIG-I E373Q mutation does not confer constitutive activity due to an exposed 2CARD module. (**A**) Small-angle X-ray scattering (SAXS) intensity curves of RIG-I and RIG-I E373Q in presence and absence of ATP. (**B**) Distance distribution functions derived from SAXS data in panel A. Calculated radii of gyration (R_g) are indicated within the legend. (**C**) Thermal shift assays in presence and absence of ATP. Melting temperatures I are indicated within the legend. (**D**) Fold change of interferon-β promoter driven luciferase activity of HEK 293T RIG-I KO cells co-transfected with a RIG-I E373Q expression vector, varying concentrations of a RIG-I Δ2CARD,E373Q expression vector and p125-luc/ pCMV-RL reporter plasmids. Firefly luciferase (FF) activities were determined in respect to Renilla luciferase (Ren) activities 24 hr after transfection. All ratios were normalized to an empty vector control. n=3, error bars represent mean values ± standard deviation. (**E**) Control Western Blot analysis of FLAG/HA-tagged constructs from panel D.

DOI: 10.7554/eLife.10859.005

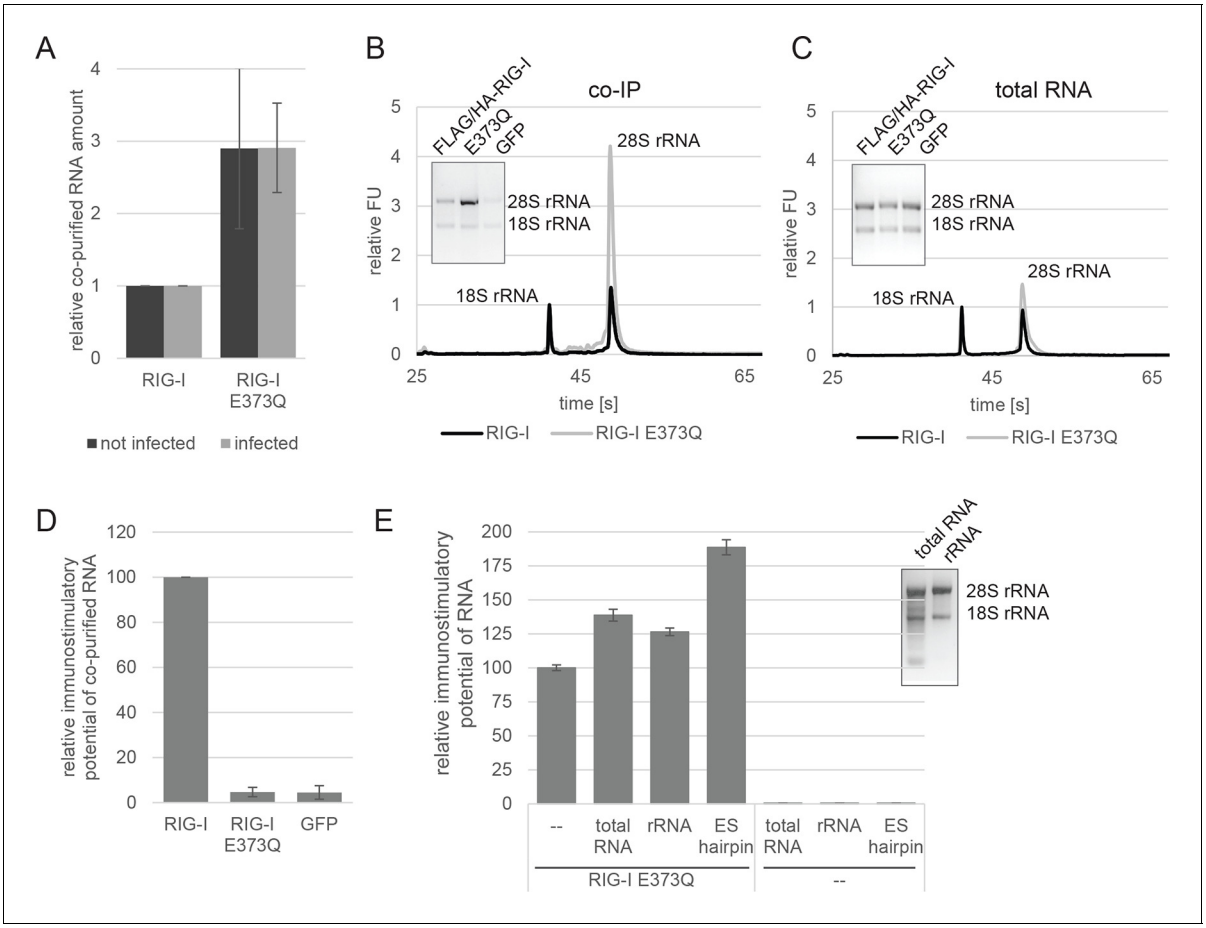


Figure 2. RIG-I ATP hydrolysis defective mutant E373Q recognizes the 60S ribosomal subunit in vivo. (**A**) Relative RNA amount co-purified with overexpressed RIG-I or RIG-I E373Q from virus infected or non-infected HEK 293T RIG-I KO cells. n=4 (infected) or n=10 (non-infected), error bars represent mean values ± standard deviation. (**B**) Bioanalyzer evaluation and agarose gel separation of RNA co-purified with overexpressed RIG-I or RIG-I E373Q from non-infected HEK 293T RIG-I KO cells. Curves are normalized in respect to 18S rRNA peaks. (**C**) Bioanalyzer evaluation and agarose gel separation of total RNA content of non-infected HEK 293T RIG-I KO cells overexpressing RIG-I or RIG-I E373Q. Curves were normalized as in panel B. (**D**) Immunostimulatory potential of co-purified RNA from RIG-I, RIG-I E373Q or GFP overexpressed in measles virus (MeV), MeV-Cko-ATU-Cs or Sendai virus Cantell (SeV) infected HEK 293T RIG-I KO cells. RNA was back-transfected into HEK 293T ISRE-FF/RFP cells together with pTK-RL transfection control. Firefly luciferase (FF) activities were determined 24 hr after transfection in respect to Renilla luciferase (Ren) activity and were normalized to the immunostimulatory potential of RIG-I associated RNA. n=4, error bars represent mean values ± standard deviation. (**E**) Immunostimulatory potential of endogenous RNA in cells overexpressing RIG-I E373Q. RNA was co-transfected into HEK 293T RIG-I KO cells together with a RIG-I E373Q expression vector and p125-luc/ pCMV-RL reporter plasmids. FF luciferase activities were determined in respect to Ren luciferase activities 24 hr after transfection. All ratios are normalized to the RIG-I E373Q control without RNA stimulation. Purified RNA was in addition analyzed on agarose gels. n=3, error bars represent mean values ± standard deviation.

DOI: 10.7554/eLife.10859.006

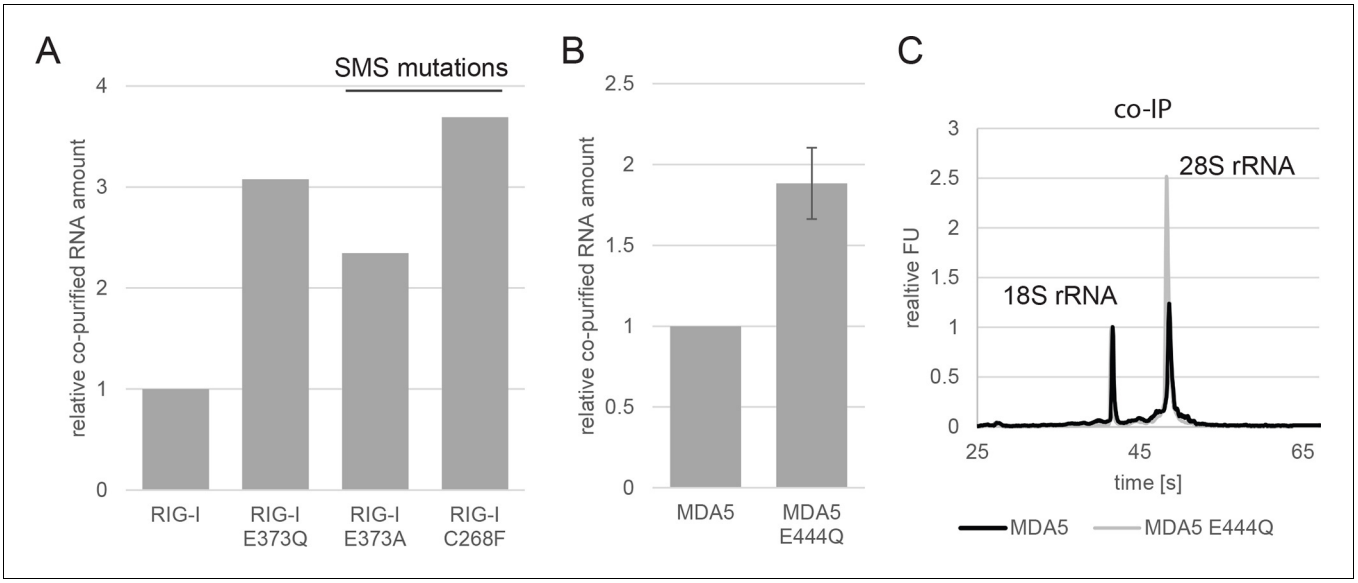


Figure 2—figure supplement 1. Analysis of RNA co-purified with RIG-I SMS or MDA5 variants. (**A**) Relative RNA amount co-purified with overexpressed RIG-I, RIG-I E373Q or RIG-I SMS variants from non-infected HEK 293T RIG-I KO cells. (**B**) Relative RNA amount co-purified with overexpressed MDA5 from non-infected HEK 293T cells. n=3, error bars represent mean values ± standard deviation. (**C**) Bioanalyzer evaluation of RNA co-purified with overexpressed MDA5 from non-infected HEK 293T cells. Curves are normalized in respect to 18S rRNA peaks.
DOI: 10.7554/eLife.10859.007

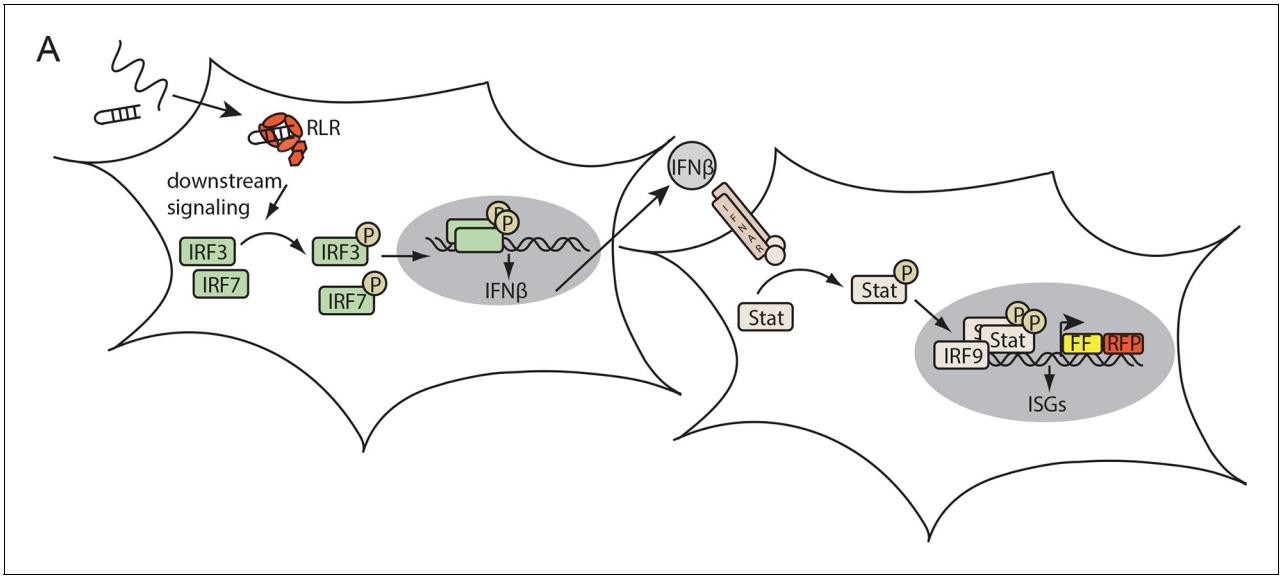


Figure 2—figure supplement 2. Assay for defining the immunostimulatory potential of different RNAs. (**A**) Endogenous RLRs in HEK 293T ISRE-FF/RFP cells (stably express firefly luciferase (FF) and RFP under control of an interferon stimulated response element (ISRE) promoter) induce a downstream signaling cascade upon binding to transfected RNA. Subsequent interferon (IFN) expression results in activation of the STAT signaling pathway which in return induces ISRE promoter driven expression of FF luciferase.

DOI: 10.7554/eLife.10859.008

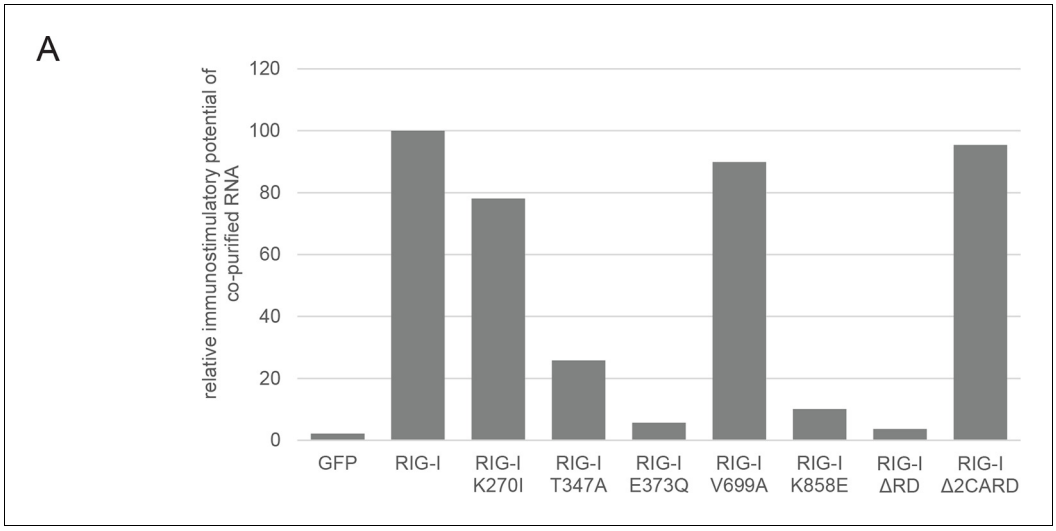


Figure 2—figure supplement 3. Immunostimulatory potential of co-purified RNA from Sendai virus Cantell (SeV) infected cells. (**A**) HEK 293T RIG-I KO cells were transfected with the indicated RIG-I mutant or GFP expression vector. RNA co-purified with the respective overexpressed protein was back-transfected into HEK 293T ISRE-FF/RFP cells (compare with ***Figure 2—figure supplement 2***). Firefly luciferase activities were determined 24 h after transfection and normalized to the RIG-I sample.

DOI: 10.7554/eLife.10859.009

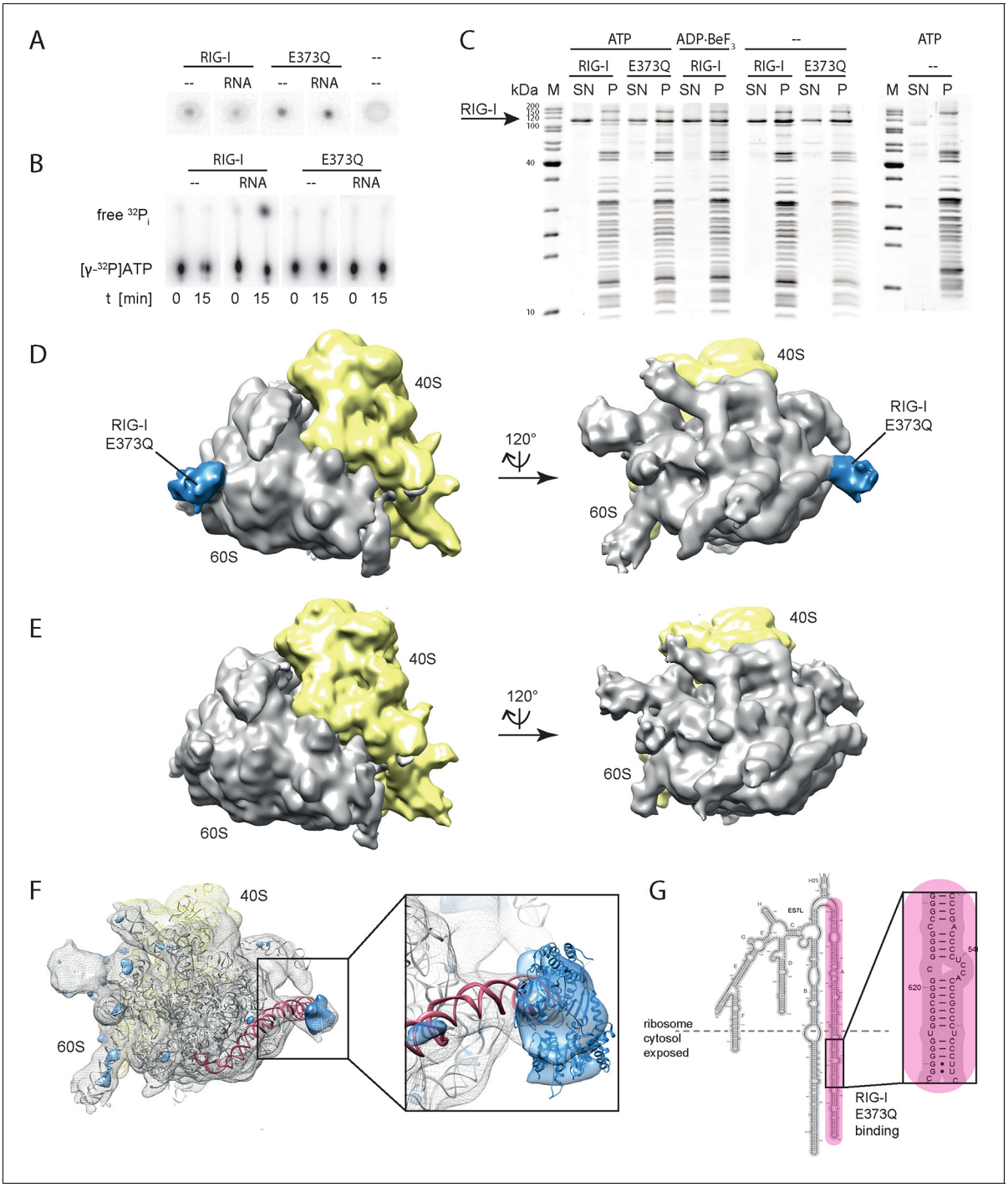


Figure 3. RIG-I ATP hydrolysis defective mutant E373Q recognizes the 60S ribosomal subunit in vitro. (**A**) DRaCALA ATP binding assay of RIG-I or RIG-I E373Q in presence or absence of RNA. (**B**) ATP hydrolysis assay of RIG-I or RIG-I E373Q in presence and absence of RNA. (**C**) Binding studies of human 80S ribosomes with RIG-I or RIG-I E373Q in presence or absence of ATP or ADP·BeF_3. Pre-formed complexes were separated on sucrose cushions via ultracentrifugation and pellet (P) as well as supernatant (SN) fractions were analyzed by SDS-PAGE. (**D**) Side views of a cryo-EM reconstruction of RIG-I E373Q (blue) bound to the human 80S ribosome (yellow: 40S subunit, gray: 60S subunit). Data was low pass-filtered at 15 Å. (**E**) Side views of a cryo-EM reconstruction of the human 80S ribosome without prior RIG-I E373Q incubation. Data filtering and color coding as in panel D. (**F**) Statistical difference map (left, σ = 2) of cryo-EM reconstructions in panels D and E reveals a significant additional density at expansion segment 7L A (ES7L-A, pink) into which RIG-I (PDB 3TMI) can be fitted (right, σ = 1.51). (**G**) Secondary structure map of the 28S rRNA ES7L (derived from (***Anger et al., 2013***) and zoom into RIG-I E373Q binding area. ES7L-A is indicated in pink (as in panel **F**).

DOI: 10.7554/eLife.10859.010

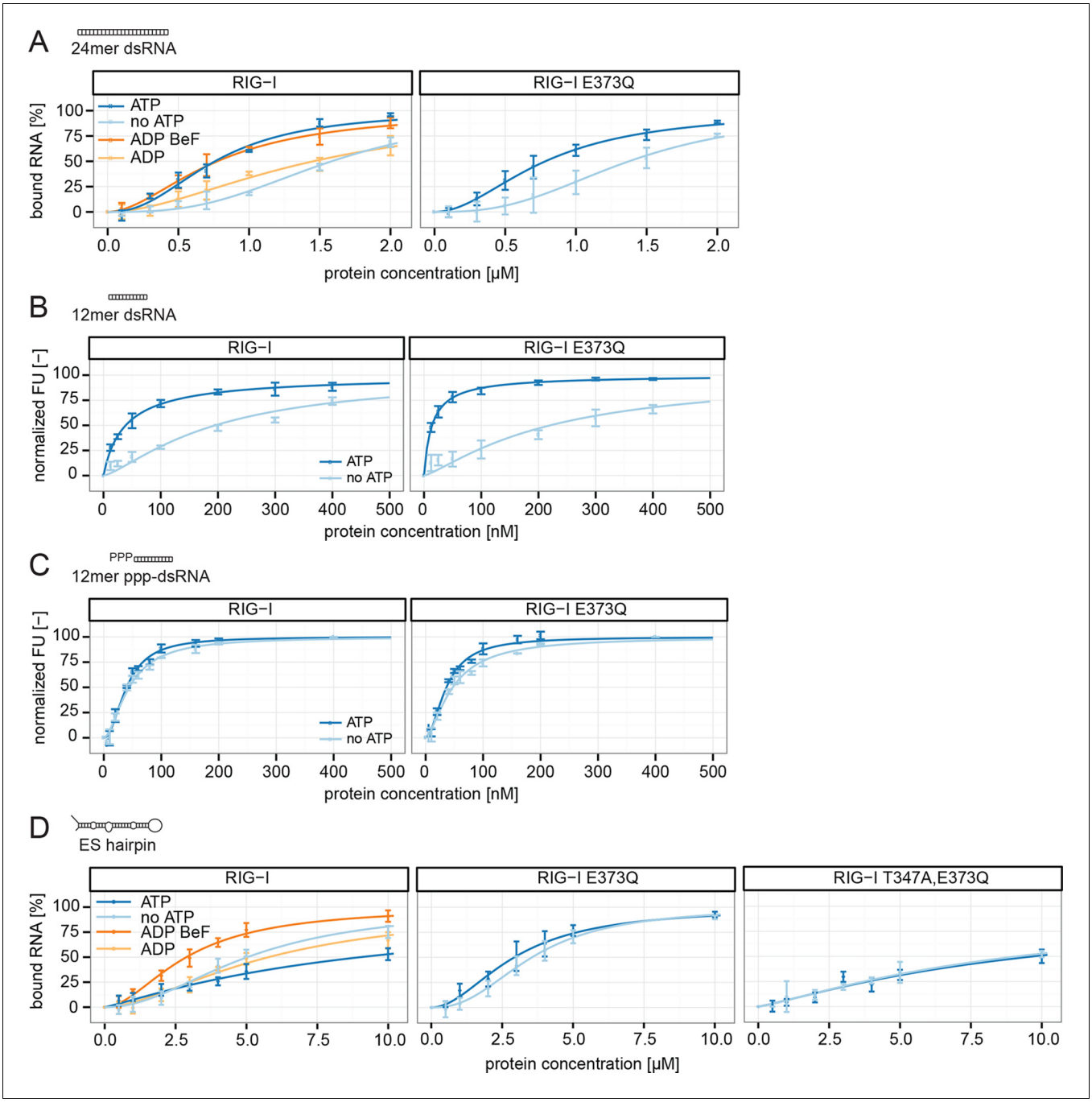


Figure 4. RIG-I's ATP hydrolysis enhances RNA end recognition and removes RIG-I from RNA stems. (**A**) Quantification of electrophoretic mobility shift assays of RIG-I or RIG-I E373Q incubated with 24mer dsRNA in presence or absence of ATP, ADP or ADP·BeF_3 (compare with ***Figure 4—Figure supplement 1B***). (**B**) Fluorescence anisotropy changes measured by titrating RIG-I or RIG-I E373Q in presence or absence of ATP into solutions containing fluorescently labeled 12mer dsRNA. (**C**) Fluorescence anisotropy changes measured by titrating RIG-I or RIG-I E373Q in presence or absence of ATP into solutions containing fluorescently labeled 12mer ppp-dsRNA. (**D**) Quantification of electrophoretic mobility shift assays of RIG-I, RIG-I E373Q or RIG-I T347A, E373Q incubated with an RNA hairpin derived from helix A of the human ribosome expansion segment 7L (ES hairpin) in presence or absence of ATP, ADP or ADP·BeF_3 (compare with ***Figure 4—Figure supplement 1C***). All binding curves were fitted using the LL.2 function of the R drc package (***Cedergreen et al., 2005***). n=3-6, error bars represent mean values ± standard deviation.

DOI: 10.7554/eLife.10859.011

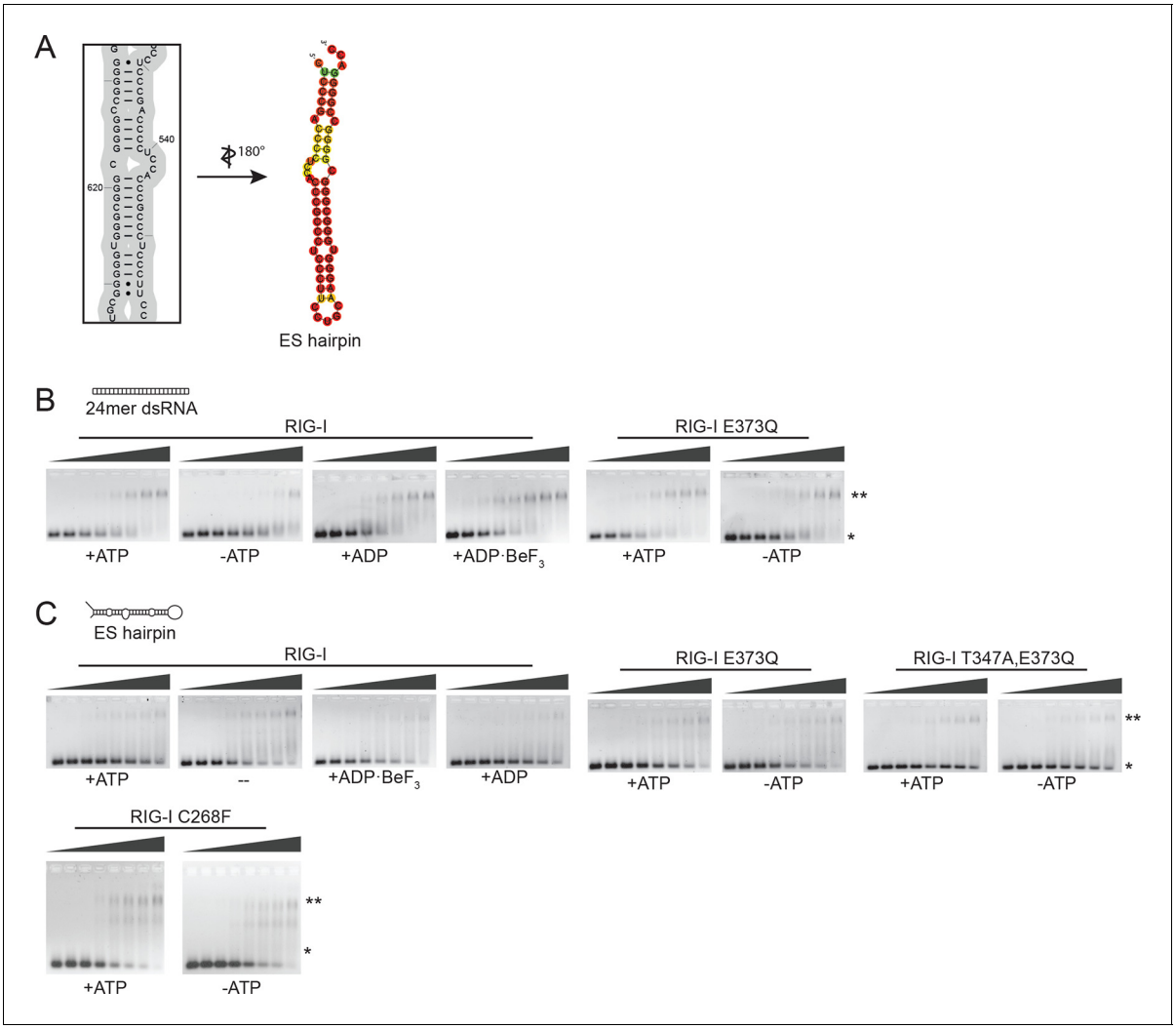


Figure 4—figure supplement 1. Design of the ribosomal expansion segment derived hairpin RNA, EMSA raw figures and control experiments with RIG-I C268F SMS mutant. (**A**) RIG-I E373Q binding site at ES7L-A was used to design a 60b hairpin RNA (ES hairpin). RNA secondary structure was determined with the RNAfold webserver (***Gruber et al., 2008***). (**B**) Electrophoretic mobility shift assays of RIG-I or RIG-I E373Q incubated with 24mer dsRNA. Complexes were pre-formed at 37 °C for 20 min, separated on agarose gels and stained with GelRed. Free RNA bands were quantified using ImageJ. Protein concentrations (from left to right): 0, 0.1 μM, 0.3 μM, 0.5 μM, 0.7 μM, 1 μM, 1.5 μM and 2 μM. *: unbound RNA, **: protein:RNA complexes. (**C**) Electrophoretic mobility shift assays of RIG-I, RIG-I E373Q or RIG-I C268F incubated with ES hairpin RNA. Complexes were pre-formed, separated and stained as in panel B. Protein concentrations (from left to right): 0, 0.5 μM, 1 μM, 2 μM, 3 μM, 4 μM, 5 μM and 10 μM. *: unbound RNA, **: protein:RNA complexes.

DOI: 10.7554/eLife.10859.012

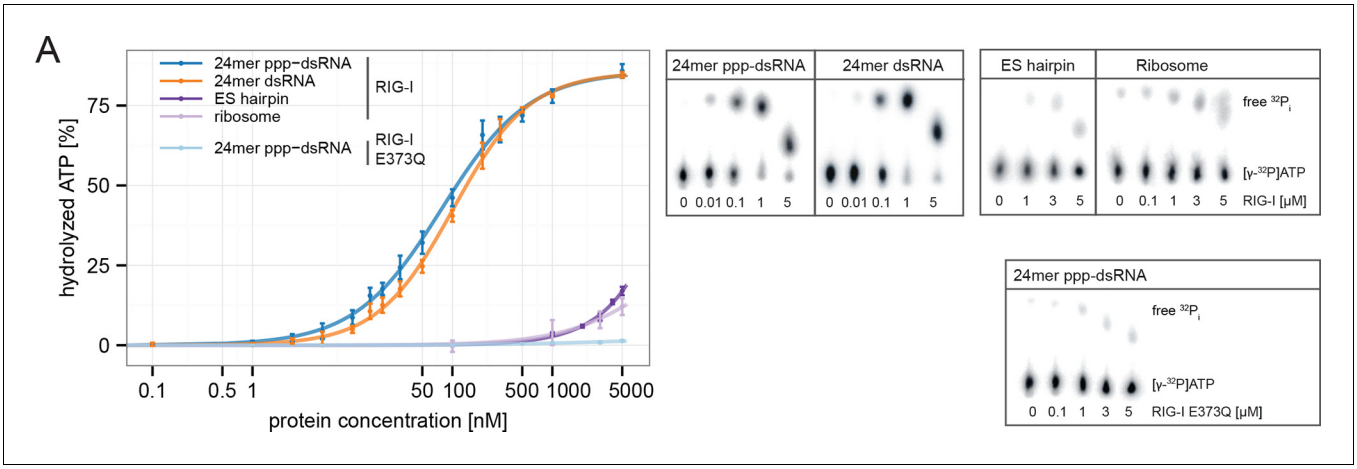


Figure 5. RIG-I's ATPase activity correlates with its RNA binding affinity. (**A**) Quantification of hydrolyzed [γ-^{32}P]ATP by RIG-I or RIG-I E373Q in presence of different RNA substrates. Reactions were allowed to proceed for 20 min at 37 °C and free phosphate was separated from ATP via thin layer chromatography. Spots corresponding to labeled ATP and labeled P_i were quantified using ImageJ. All curves were fitted using the LL.2 function of the R drc package. n=3, error bars represent mean values ± standard deviation.

DOI: 10.7554/eLife.10859.013

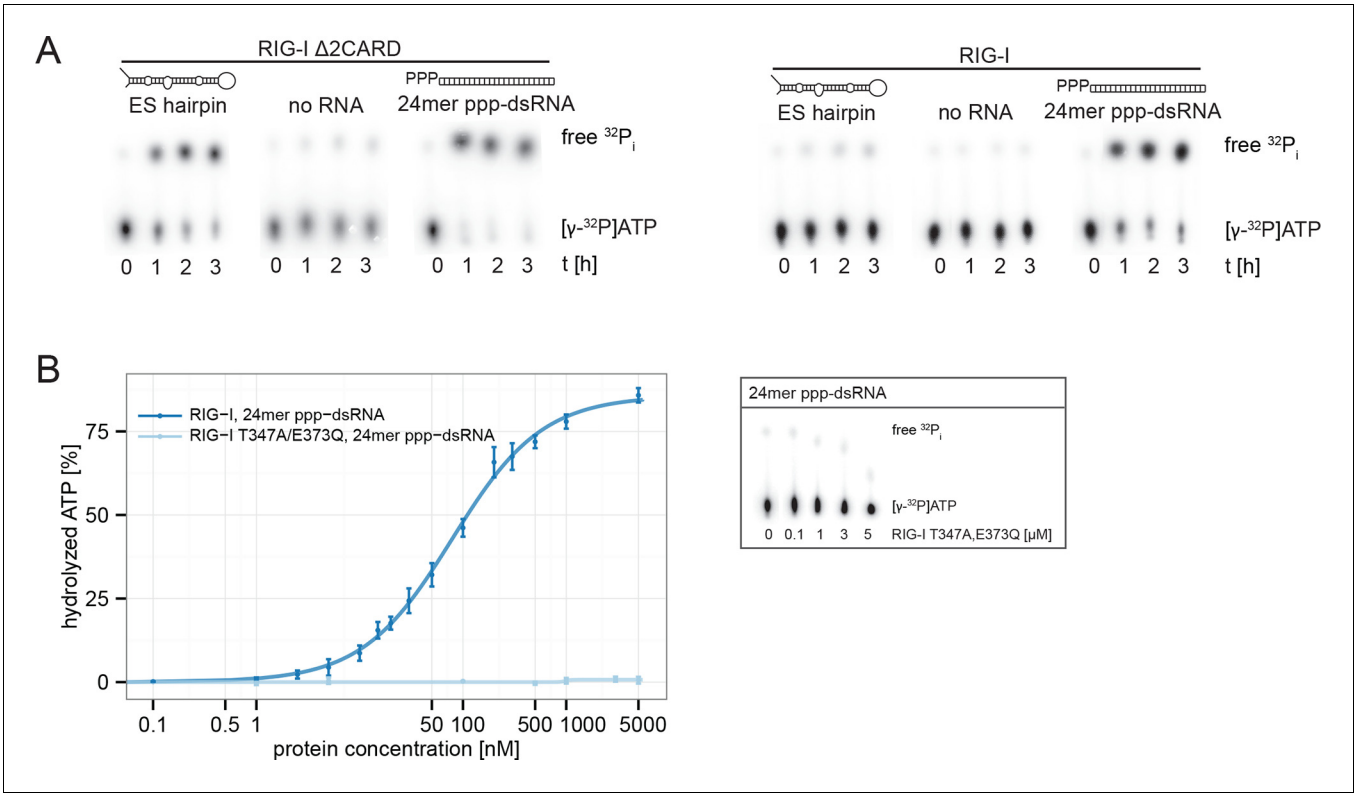


Figure 5—figure supplement 1. RIG-I's 2CARD module reduces the ATP hydrolysis activity. (**A**) Measurement of ES hairpin or ppp-dsRNA stimulated [γ-^{32}P]ATP hydrolysis of RIG-I or RIG-I Δ2CARD. Reactions were monitored over 3 hr at room temperature and free phosphate was separated from ATP via thin layer chromatography. (**B**) Quantification of hydrolyzed [γ-^{32}P]ATP by RIG-I T347A,E373Q in presence of 24mer ppp-dsRNA. Reactions were allowed to proceed for 20 min at 37 °C and free phosphate was separated from ATP via thin layer chromatography. Spots corresponding to labeled ATP and labeled P_i were quantified using ImageJ. Curves were fitted using the LL.2 function of the R drc package. n=3, error bars represent mean values ± standard deviation.

DOI: 10.7554/eLife.10859.014

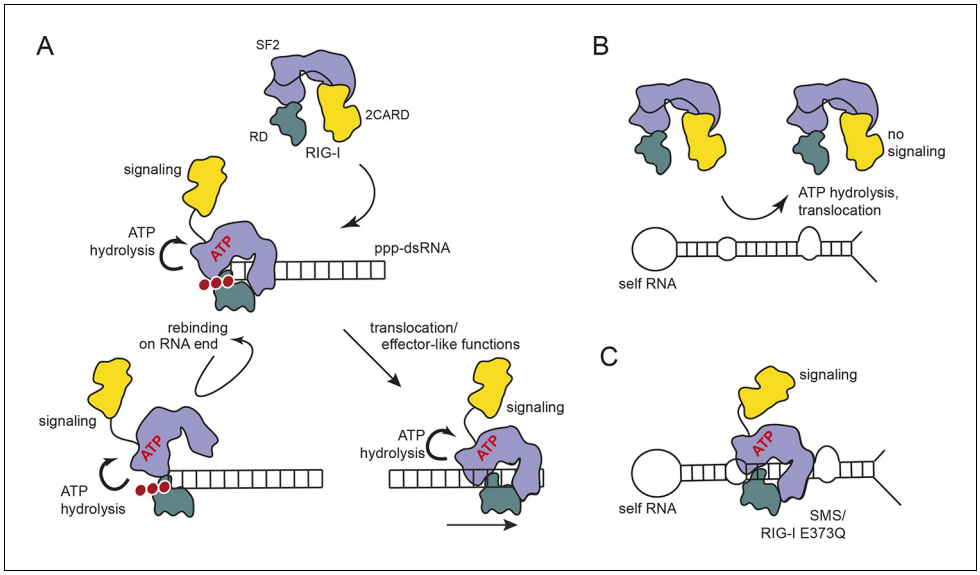


Figure 6. Proposed model for impact of ATP on RIG-I signaling on different RNAs. (**A**) RIG-I recognizes tri- or diphosphorylated double-stranded RNA and preferentially binds to the RNA end through its regulatory domain (RD, green). Binding of ATP-SF2 (purple) to the dsRNA releases the 2CARD module (yellow) and activates the downstream signaling process. ATP hydrolysis displaces the SF2 domain from dsRNA leading to either rebinding at the RNA end (tethered by RD) or to translocation along the RNA. (**B**) In healthy cells, sustained binding of RIG-I to self-RNA containing dsRNA stretches is prevented by ATP hydrolysis. The SF2 domain can be sufficiently displaced because the RD does not provide a high affinity tether. (**C**) Mutations that allow ATP promoted binding of dsRNA and displacement of the 2CARD module, but prevent ATP hydrolysis dependent dissociation of SF2 from dsRNA, such as those underlying atypical Singleton-Merten Syndrome, will result in an unintended signaling through self-RNA.

DOI: 10.7554/eLife.10859.015

PUBLICATIONS

3 Discussion

3.1 cGAS *IN VIVO* LIGAND - LENGTH VERSUS SEQUENCE

Despite extensive structural studies on dsDNA recognition by cGAS, several questions remain open: the mode of binding of cGAS to long immunostimulatory DNA (ISD), how short DNA can activate cGAS in a sequence-specific way, how cGAS can act as a major sensor for DNA despite a comparatively low affinity for DNA along with a very promiscuous recognition and which is the nature of cGAS endogenous ligands arising from cell damage under stress conditions. Structures of cGAS bound to different dsDNA along with *in vitro* assays do not comply with results of cGAS activation performed in cell assays (106,109,135). In the crystal structure a minimum of 16 bp are required in order for the DNA to span both binding sites of cGAS in the 2:2 cGAS-DNA complex (109). In agreement with this, 12 bp DNA does not activate mouse cGAS efficiently *in vitro*, while 18 bp evidenced about 90% activity in comparison to salmon sperm DNA (109). However, DNA length dependence (> 40 bp) for interferon stimulation by cGAS *in vivo* is not readily explained by the 2:2 complex structures of cGAS binding to two short dsDNAs. In fact, the 2:2 complex structures even indicate that the binding of long dsDNA by cGAS is not sterically favorable (FIG. 18) (109). Binding of cGAS to internal sites on long dsDNA would lead to collision of both DNA duplexes, whereas binding of cGAS to DNA ends avoids these potential clashes and therefore seems to be at first more plausible.

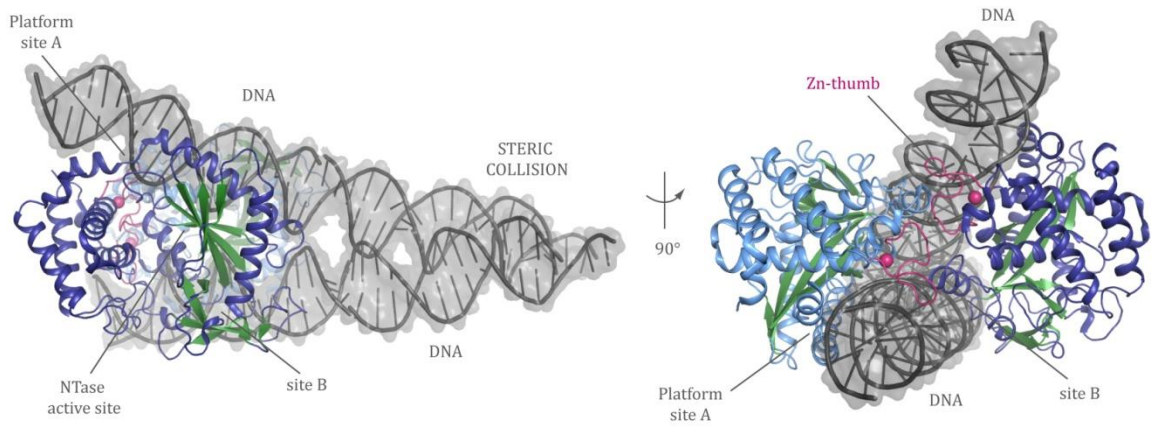


FIGURE 18: MODEL OF 37 bp DNA CLASHING IN THE cGAS OLIGOMER

Top and front views of the MAB21 domain of cGAS in a 2:2 cGAS:dsDNA complex (PDB: 4LEY) modeled with 37 bp long DNA. Both 37 bp DNA strands are depicted in grey, cGAS protomers in light and dark blue respectively and the Zn-thumbs in pink. The binding mode of cGAS to the two 37 bp long dsDNA leads to a clash between both strands. Thus, DNA length dependence (minimum 40 bp) for interferon stimulation by cGAS *in vivo* cannot be explained by the current model.

Nevertheless this does not explain how cGAS detects and measures the length of the DNA duplex for activation *in vivo*. Moreover, analytical ultracentrifugation experiments detected higher oligomeric species of cGAS bound to dsDNA *in vitro*, even though in probably non-physiological cGAS concentrations. Further, in the available crystal structure each DNA

duplex stacks end-to-end with another duplex mimicking a continuous 36 bp DNA with cGAS dimers bound along the DNA duplex axis (109). Taking into account how other sensors, such as MDA5 sense the length of the bound nucleic acid, it is tempting to speculate that cGAS might bind DNA co-operatively and may even dimerize along the length of the DNA duplex for this purpose.

cGAS is a generic DNA sensor and is thought to be relatively unspecific with its main requirements being double-strandedness and a sufficient length over 40 bp (135). Recent experiments revealed that cGAS is activated in a structure- and sequence-dependent manner by a minimal immunostimulatory DNA motif (221). This so-called 'strong-stop DNA' or G-YSD motif was detected in HIV-1 early transcripts and endogenous retroviral elements. Hence, DNA recognition by cGAS was shown to be more specific than previously thought and first evidence was provided that single-stranded DNA when containing guanosine-rich regions and a short duplex stem loop structure is sufficient to effectively trigger cGAS activation. Another possibility is that the polyglutamine binding protein 1 (PQBP1) mediates cGAS specificity for retroviral G-YSD motifs. PQBP1 was demonstrated to serve as a specific co-receptor directly binding reverse-transcribed viral cDNA and to interact with cGAS in order to initiate the immune response against retroviruses (222). Besides retroviral DNA, PQBP1 is not required for cGAS-dependent immune response to cytosolic dsDNA. It is still not clear if cGAS requires other co-sensors analogous to PQBP1 in order to detect other kinds of cytoplasmic dsDNA. Regarding the reported affinity of cGAS (MAB21 domain) for generic dsDNA and its broad range of specificity, one could suspect that either other co-sensors or the so far uncharacterized N-terminal unstructured region of cGAS contribute to its particular mode of PAMP recognition and subsequent activation (109,135,136).

cGAS has been shown to recognize mislocalized dsDNA, such as mitochondrial DNA that is released into the cytosol after herpesvirus induced stress (117). The nature of other potential endogenous ligands for cGAS has to be uncovered, since aberrant cytoplasmic localization of DNA is recognized by cGAS leading to a robust interferon response to self nucleic acids and therefore autoimmunity (223).

3.2 STRUCTURE-BASED ANALYSIS OF ANCIENT cGAS ORIGIN AND OTHER NUCLEOTIDYLTRANSFERASES

Structural studies performed in this work have a drastic impact on our understanding of the evolutionary origin regarding the template-independent nucleotidyltransferases cGAS and OAS. This conserved family of nucleic acid sensors in innate immunity seemed to have diverged from a common 3´-specific ancestor right at the beginning of metazoan evolution (224). OAS proteins and cGAS developed the unique feature to form 2´-5´-linked second messengers and both proteins belong to the large and diverse family of nucleotidyltransferase fold proteins. OAS proteins were placed in the well characterized class of template-independent polymerases due to structural conservation with the 3´-specific poly

(A) polymerase (PAP), whereas cGAS was classified as a member of the MAB21 subfamily (150). Poly (A) polymerases are involved in mRNA metabolism by adding a poly-A tail to mRNA (225). Conservation among nucleotidyltransferases is focused around the phosphate binding loop (P-loop) and the first two divalent ion coordinating catalytic carboxylate residues. The classical 5-stranded β-sheet domain (palm domain – hand analogy used by Steitz and others (140)) on the N-terminal lobe harboring the catalytic triad is found in all members of the nucleotidyltransferase family, while the thumbs and finger domains as found in template-dependent polymerases are replaced in PAP, OAS proteins and MAB21 family members by a C-terminal lobe composed of a four helix bundle (192).

Main aim of this work was to decipher the evolutionary path within members of the MAB21 family and cGAS in order to improve the enzymatic understanding of this newly discovered subfamily of nucleotidyltransferase fold proteins. cGAS is present in choanoflagellates (*Monosiga brevicollis*) and metazoans as early as cnidarians (*Nematostella vectensis*) while completely missing in nematodes (226). MAB21 proteins were first discovered in the nematode (*Caenorhabditis elegans*), are present in all metazoa and in contrast to cGAS are missing in the choanoflagellates (171). MAB21 proteins seem to have evolved earlier in evolution as cGAS. In some species such as *H. magnipapillata*, *T. castaneum*, *Drosophila virilis*, *Drosophila persimilis*, *Drosophila pseudoobscura*, cephalochordate *B. florida* and *Danio rerio* two cGAS homologs were identified (226). MAB21 as well duplicated once in the vertebrate genome (MAB21L1 and MAB21l2) and in *Drosophila* similarly, two MAB21 homologs were mapped (171). In early vertebrate evolution, cGAS developed some key features: the zinc-thumb, which is missing in all invertebrate cGAS variants and the N-terminal extension, which consists of over 150 amino acids length in vertebrates whereas in cnidarians and insect it is as short as 70 residues (226). MAB21 proteins do not possess special features except the common MAB21-domain fold and an extensive positively charged surface.

Residues orienting the donor substrate in the active site (MAB21L1 – Y272, K275, K255) are conserved in 3′-specific nucleotidyltransferases, such as the poly(A) polymerase and CCA-adding enzyme (CCA-adding enzyme are responsible for maturation or repair of the functional 3′-end of tRNA by adding the CCA tail (227)) along with cGAS, OAS proteins and MAB21 proteins (146). Different reaction specificities are achieved by variation of active site residues on the acceptor side (N-terminal lobe). Critical catalytic residues in the N-terminal lobe of human cGAS (G212, S213, E225, D227 and D319) were found to be conserved in all cGAS species, OAS1 and even the bacterial nucleotidyltransferase DncV (226). DncV was identified as a virulence factor on pandemic island-1 of *Vibrio cholerae*, which is capable of synthesizing a mixed base cyclic GMP-AMP molecule with 3′-5′ phosphodiester linkages (3′-5′ cGAMP) (228). Structural characterization of DncV revealed that DncV is a homolog of cGAS (229-232). While previously described bacterial cyclic dinucleotide synthases, such as DisA, formed oligomers where each of the two phosphodiester bonds was generated simultaneously in active sites of adjacent monomers, DncV resembles cGAS performing the same reaction with a single active site (142,229). Unlike cGAS, DncV is in a constitutively active conformation evidencing a helix break of the ‘spine’ with L24 already oriented toward the catalytic pocket and no requirement for previous arrangement of the catalytic triad (229).

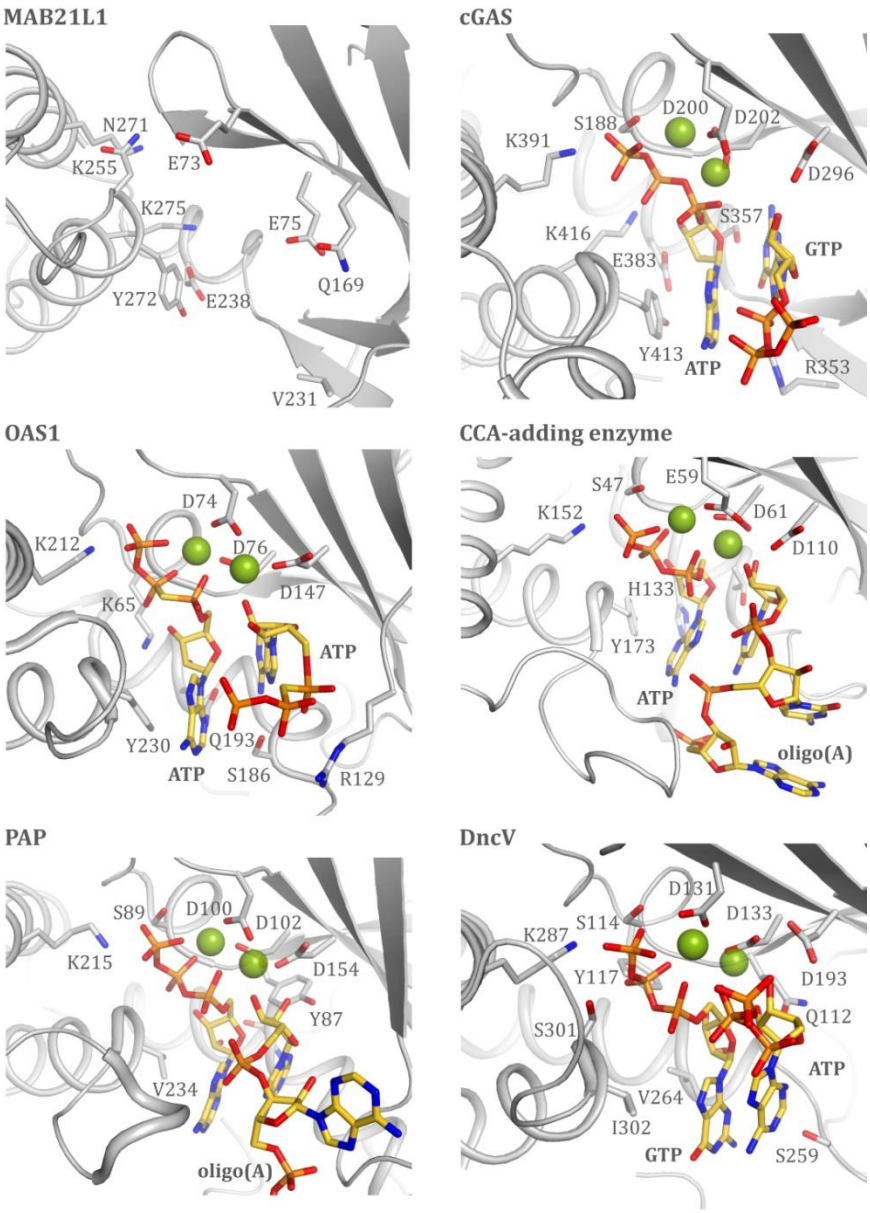


FIGURE 19: ACTIVE SITES OF DIFFERENT NUCLEOTIDYLTRANSFERASE FOLD PROTEINS

Close-up views of the NTase active sites of human MAB21L1 (PDB: 5EOM), porcine cGAS (PDB: 4KB6), porcine OAS1 (PDB: 4RWN), *Archaeoglobus fulgidus* CCA-adding enzyme (PDB: 1TFW), *Saccharomyces cerevisiae* PAP (PDB: 2Q66) and *Vibrio cholerae* DncV (PDB: 4UOM) shown in the same orientation together with the respective substrates in yellow. Residues responsible for Mg^{2+} ion coordination and implicated in catalysis are annotated. For illustrative reasons missing Mg^{2+} ions in the crystal structures were added and mutant residues changed to wild-type residues.

Strikingly, despite no requirement for activation by a oligonucleotide as its counterpart cGAS, DncV retains a long basic cleft with conserved residues along the surface corresponding to the 'platform' in cGAS and OAS (FIG. 20) (231). MAB21 proteins as well have in common the extensive positively charged surface suggesting possible involvement of interactions with oligonucleotides.

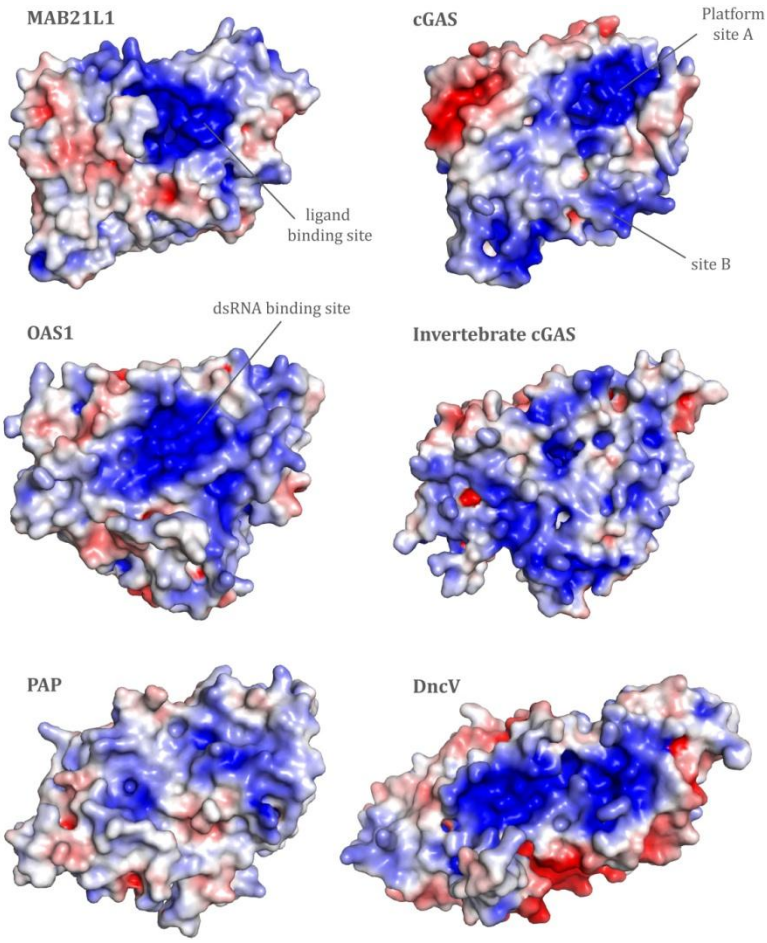


FIGURE 20: COMPARISON OF SOLVENT ACCESSIBLE ELECTROSTATIC SURFACE OF NTases

Solvent accessible electrostatic surface representations of human MAB21L1 (PDB: 5EOM), human cGAS (PDB: 4KM5), human OAS1 (PDB: 4IG8), model of invertebrate cGAS from *Nematostella vectensis*, *Saccharomyces cerevisiae* PAP (PDB: 2Q66) and *Vibrio cholerae* DncV (PDB: 4UOM) colored by charge (blue = 5kT/e to red = -5kT/e). The ligand binding pocket in MAB21L1 and the dsDNA and dsRNA binding sites in cGAS and OAS1 respectively are indicated with an arrow.

2´-specificity of cGAS and OAS proteins results from a different orientation of the ribose (i.e. the *syn* and *anti* conformations) of the acceptor substrate where the 2´-OH, and not the 3´-OH, is positioned near the phosphate of the donor nucleotide for catalysis (229). cGAS reprogramming experiments relying on the bacterial DncV structure showed that the mutation R353I is enough for cGAS to produce the bacterial 3'-5'-cGAMP (229). These distinct recognition elements enable cGAS to distinguish between guanine and adenine in order to bind 2'-5'-cGAMP in a preferred orientation. DncV has a more promiscuous active site that catalyses the reaction of cGAMP in opposite direction to cGAS and is even capable of producing other types of CDNs such as c-di-GMP and c-di-AMP (228,230). Notably, these residues, which confer cGAS specificity for the guanine base, are missing in MAB21L1. MAB21 proteins only possess two out of these five catalytic NTase core residues (E73 and E75) (226). Therefore, MAB21L1 probably has less specificity for a possible acceptor substrate, than the evolutionary more developed nucleotidyltransferase cGAS. Besides being in an

inactive conformation and partially lacking the catalytic triad (E73, E75 and Q169), MAB21L1 still possesses conserved features of a possibly active nucleotidyltransferase. The C-terminal α-helical bundle is rigid and residues responsible for the donor nucleotide coordination, such as ATP, are conserved in MAB21L1. The N-terminal lobe is only partially conserved and orientation of the catalytic residues on the flexible β-sheet bundle varies depending on major conformational changes. MAB21 protein residues on the NB-loop participating in catalysis could not be assigned, due to the lack of a MAB21L1 structure in an active conformation.

3.3 THE SEARCH FOR THE MAB21 PROTEIN FUNCTION

Apart from sharing a high degree of structural homology, MAB21 proteins and cGAS seem to have developed clear distinct functions in the organism. Previous cell localization studies with MAB21 proteins showed that MAB21 can be found in the nucleus as well as in the cytoplasm (191,194). Vertebrate cGAS however, was compartmentalized to the cytosol where it evolved into the major and nonredundant cytosolic DNA sensor (103). dsDNA binding characteristics of vertebrate cGAS are lacking in the invertebrate version of cGAS (233). While vertebrate cGAS homologs can be confidently identified based on conserved catalytic residues and DNA binding properties, invertebrate cGAS proteins are likely to be orthologs of human MAB21L2. According to the phylogenetic tree, cGAS is suggested to have originated during the transition from invertebrates to vertebrates (FIG. 21) (1).

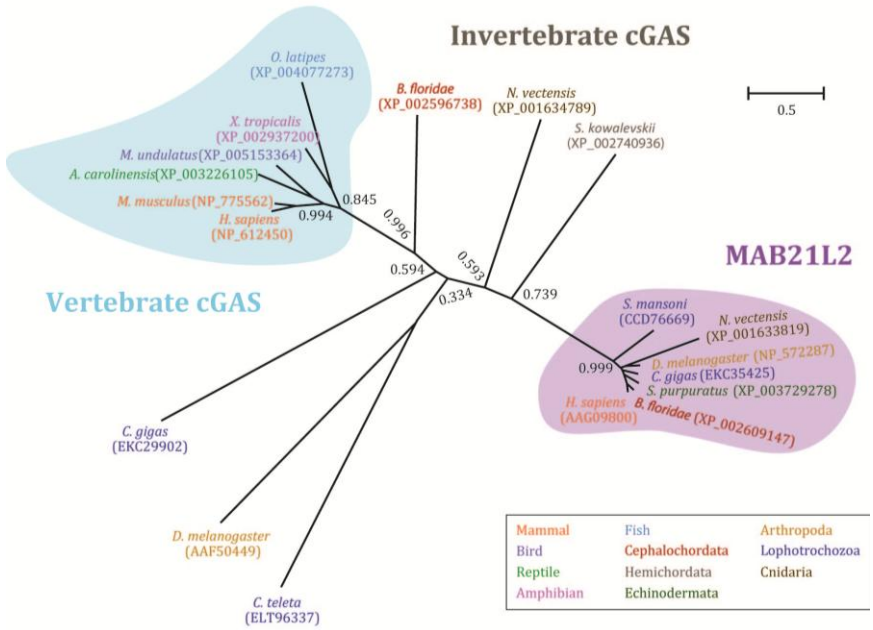


FIGURE 21: PHYLOGENETIC ANALYIS OF cGAS AND MAB21L2

cGAS and MAB21L2 proteins from representative species were obtained by NCBI BlastP search. Invertebrate cGAS proteins are closer to be orthologs of human protein MAB21L2. This figure was adapted from (1).

The Toll-like Receptors have an evolutionary ancestor in *Drosophila* - the Toll protein, which exhibits the dual role of immune signaling and dorso-ventral patterning of the embryo (11,12). Likewise, the MAB21-like proteins are known to be responsible for the formation of the dorso-ventral axis and may have served as ancestors for the pathogen recognition receptor cGAS (189).

Recently, functional cGAS and STING homologs were experimentally identified in the sea anemone *Nematostella vectensis* (233). Anemone cGAS was able to induce STING activation in cells, in contrast to the MAB21 homolog in *N. vectensis* (233). This indicates that CDN synthesis for STING activation is not a feature exclusively acquired by vertebrate cGAS. Moreover, phosphodiesterase assays in cells with *Vibrio cholera* cGAP1, which efficiently hydrolyzes exclusively 3´, 3´cGAMP, suggest that the CDN synthesized by invertebrate cGAS contains only 3´-5´-phosphodiester linkages (233,234). However, recombinant purified anemone cGAS showed no detectable activity *in vitro* even in the presence of dsDNA (233). These results suggests that anemone cGAS requires activation by a ligand present within the cellular environment other than generic dsDNA as in the case of vertebrate cGAS. MAB21 proteins likewise developed an affinity for oligonulceotides, which is lacking in other MAB21 family members (such as MB21D2, data not shown). Due to their less confined localization, it seems plausible that the oligonucleotide ligand has to have specific features to be recognized by MAB21 proteins, in contrast to the generic dsDNA activating vertebrate cGAS. It is still unclear, what the chemical properties of CDNs are that invertebrate cGAS is capable of synthesizing and all together what the previous role of cGAS was before it evolved to a pathogen recognition receptor.

This work provides alongside unique structural information on MAB21L1, some valuable insights required to unravel the molecular function of MAB21-like proteins. The structure of MAB21L1 bound to CTP revealed a potential ligand binding site. This positively charged surface corresponds to the dsDNA binding 'platform' of cGAS, the dsRNA binding site in OAS1 and recently DncV activity was shown to be regulated by binding of 5-methyltetrahydrofolate diglutamate to a similar ligand binding pocket (106,135,138,232). CTP is unlikely to be the physiological ligand, since it was not enough in order to trigger conformational changes of MAB21L1 required for activation. Instead, CTP or ATP, which was bound by MAB21L1 as well, could represent part of a physiological ligand or mimic the chemical structure of a nucleotide-related molecule or a specific oligonucleotide. Since MAB21L1 displayed the highest affinity for ssRNA *in vitro* and a binding preference for a tri-phosphate, it is tempting to speculate about binding to, e.g. the cap-structure of an mRNA or microRNA precursor as MAB21L1 was also detected in the nucleus (191,194).

Besides having most of the characteristics common for a nucleotidyltransferase, no such kind of activity was detected for MAB21-like proteins yet. Similar, MiD51 and MiD49, members of the MAB21 subfamily, were demonstrated to execute distinct functions in the cell, despite of harboring the nucleotidyltransferase fold. Binding of ADP to the NTase active site was shown to stabilize the interactions between Drp1 and MiD51 (161,162). The MiD51 active site resembles the one in proteins NF45 and NF90, which are known to be involved in a variety of

cellular processes regulating gene expression on a transcriptional as well as translational level (161,235). These nucleotidyltransferase fold proteins lost crucial catalytic residues during evolution leading to a lack of activity (236). Thus, NF45 and NF90 exert their functions via protein-protein and protein-RNA interactions (236,237). Another example is OASL, a unique member of the OAS family, which also lacks NTase activity and fulfills its function via a C-terminal domain homologous to ubiquitin (93).

Several nucleotidyltransferase fold proteins are known to oligomerize making use of different interfaces in order to exert their cellular functions. MiD51 dimerization is mediated by the N-terminal α-helical segments, whereas NF45 and NF90 dimerize via their C-terminal α-helices and an N-terminal extension of NF45 further contributes to the dimerization interface (162,236). cGAS instead, was demonstrated to require ligand binding in order to oligomerize (109,139). Data obtained in the course of this work show that, MAB21L1 forms a highly symmetric double-pentameric oligomer in the crystal, which however could not be detected in solution. MAB21L1 might require ligand binding for oligomerization as in the case of cGAS, since MAB21L1 alone did not readily oligomerize in solution. Since oligomerization is a key feature of several nucleotidyltransferases, it is plausible to suggest that this oligomeric state of MAB21L1 could have a physiological relevance.

Regarding the pathophysiology of MAB21-like proteins in human disease, single point mutations E49K, R51C/H and R247Q in MAB21L2 were found in patients with eye malformations termed ocular coloboma (191). Ocular coloboma describes the spectrum of eye malformations, including microphtalmia resulting from failure of optic fissure closure during embryonic eye development (optic fissure closure defect - OFCD) (194). Patients suffering from OFCD normally have additional systemic anomalies, most commonly of neurological, muscoskeletal and facial, urological and genital, or cardiac nature (238). The role of MAB21-like proteins in embryonic eye development is largely unknown. The most common genetic causes of structural eye malformations are both involving retinoic acid metabolism and transport or dosage-sensitive transcription factors, such as *sox2, otx2* or *pax6* (184,239-241). The latter was demonstrated to regulate MAB21 expression levels. Since the molecular function of MAB21 proteins remains elusive, the structural information provided in this work was used to explain the phenotype caused by the patient-related mutations. The mutations E49K and R51C/H were demonstrated to affect protein stability and to disrupt the fold of the surface accessible loop β4-α4. The same loop was demonstrated to play an important role in protein-protein interactions between MiD51 and Drp1 (161,162). As in this case, MAB21-like proteins loop β4-α4 could interact with possible still unknown interaction partners and lack of proper function due to destabilization of this loop could then be responsible for the drastic phenotype.

As common for other genes with early embryonic expression, knockout studies of MAB21-like proteins caused embryonic lethality (183). In addition, redundancy in the *mab21* gene family masks several of the early effects of *mab21* gene mutation and only uncovers the role of these genes in later stages of development (188). Nevertheless, acquired information from several studies in different model organisms and for the three members MAB21L1, MAB21L2

and MAB21L3 gives first insights into their crucial role in development. Recently, several reports revealed that molecular mechanisms controlling the differentiation of multiciliated cells are strictly conserved in mammalian upper airways and in the *Xenopus* embryonic epidermis (242,243). Since human airway epithelial cells (HAECs), the main tool for investigation of MCE biology are technically demanding, *Xenopus* revealed to be a powerful and valuable model to address the poorly understood vertebrate MCE biology (244). Important information regarding cilia specification and pathways involved was retrieved from these recent findings. Crucial links regarding the function of MAB21-like proteins in this field are presented in this work.

Fine-tuning of BMP signaling activity was demonstrated to be required for the correct organogenesis of the *Xenopus* embryonic epidermis (245). Specification of multiciliated cells (MCCs), ionocytes and small secretory cells (SSCs) requires the attenuation of BMP activity, identifying the BMP pathway as a key regulator of vertebrate mucociliary epithelium differentiation and intercalation. Previous experiments revealed MAB21L2 to modulate and antagonize BMP signaling through direct interactions with its effector transcription factor SMAD1 (188). On top of that, MAB21L3 was recently demonstrated to act downstream of the Notch pathway and act upstream of the master regulator *foxj1* and *multicilin* in cilia formation (190). The BMP and Notch pathways are linked in the *Xenopus* epidermis controlling cell fate choices. BMP4 injection results in drastic decrease amongst others of *foxj1* expression levels in MCCs and a strong upregulation of the Notch ligand *delta-like-1* (245). Likewise blocking of the Notch pathway increases the number of MCCs (246,247). Both pathways seem to have an inhibitory effect on cells preventing them to engage in cell fate choices to form MCCs. MAB21L3, which is downregulated by Notch, counteracts by up-regulating the master genes for MCC and ionocyte cell fate specification (FIG. 22) (190).

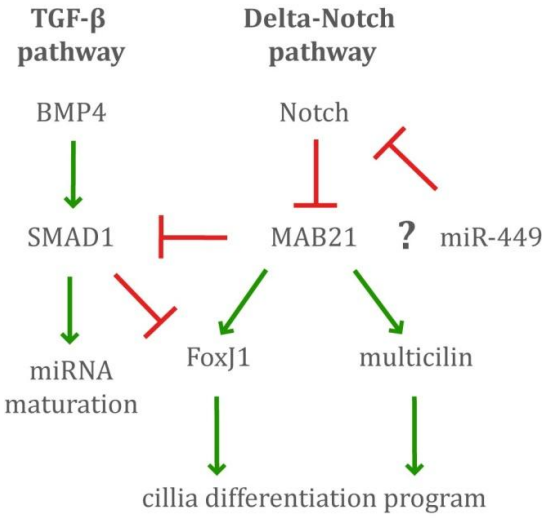


FIGURE 22: INVOLVEMENT OF MAB21-LIKE PROTEINS IN CILIOGENESIS

MAB21-like proteins are involved in the regulation of cell fate in the context of ciliogenesis. MAB21L3 was shown to be counteracted by Notch signaling and MAB21L2 was demonstrated to antagonize the effects of BMP4 by interacting with SMAD1. While MAB21L3 is the major regulator for cilia differentiation, SMAD1 was shown to inhibit cilia development. The Notch signaling is in turn regulated by miR-449.

Another factor known to regulate MCC differentiation trough interference with the Notch pathway in *Xenopus* embryonic epidermis and HAECs is the microRNA family miR-449. The miR-34/ 449 family consists of six homologous highly conserved miRNAs at three genomic loci (miRNA-34a, miRNA-34b, miRNA-34c, miRNA-449a, miRNA-449b and miRNA-449c) (248). miR-449 are the most strongly induced microRNAs during epithelium differentiation and therefore conserved key regulators of vertebrate multiciliogenesis (249). Differentiation of ciliated cell progenitors requires inhibition of the Notch pathway by miR-449. Accordingly, conserved miR-449 binding sites were found in the 3´-untranslated region (3´-UTR) of *delta-like-1* and *notch* (250). Experiments with mouse and *Xenopus* MCCs deficient in all miR-34/ 449 evidenced a decrease in cilia length and number caused by defective basal body maturation and apical docking (249). A similar phenotype with decreased number of MCCs was detected when *mab21l3* was knocked down with an antisense morpholino (190). Therefore MAB21-like proteins seem to operate in concert with miR-449 in cell fate specifications of MCCs. miR-449 were detected almost exclusively in tissues harboring high amounts of cilia, such as the brain, reproductive organs and respiratory tract (251,252). These partially overlap with tissues where high mRNA levels of MAB21-like proteins were measured (172-174,176). It remains to be shown, if miR-449 are also present in non-motile cilia, such as the connecting cilia of photoreceptors in the retina where for instance MAB21L1 has high expression levels (178). Another link represents the requirement of normal ciliogenesis for correct neural tube closure connecting cilia with brain development (253-255). MAB21-like proteins were shown to be involved in ciliogenesis as well as brain development. Previously studies where XMAB21L2 antisense RNA was injected blocked closure of the neural tube (174).

How large amounts of miR-449 expression are induced and regulated during cilia specification are still largely unknown. Transcription factors responsible for mucociliary epithelia, such as Foxj1, could be involved. MAB21-like proteins, which regulate Foxj1, could represent another possible candidate (190). MAB21-like proteins might regulate miRNA via another mechanism. For instance MAB21L2 was shown to antagonize BMP4 and to directly bind SMAD1 (188). Maturation of pri-miRNA to pre-miRNA by the specific nucleases Drosha and Pasha is promoted by SMAD1 and BMP signaling (256,257). If this is disrupted by MAB21-like proteins, miRNA precursors are not cleaved and transported into the cytosol for further processing through the ribonuclease Dicer.

3.4 RLRs ATPase AND TRANSLOCASE ACTIVITY

As a strategy to distinguish between self vs. non-self nucleic acids, RLRs developed at least two different levels of control limiting RLR mediated signaling to viral non-self RNA. First, self-RNA editing by ADAR1 and methylation, modify particular types of endogenous RNA prone to serve as ligands for RIG-I and MDA5 in such a way that unintentional recognition is impeded (42,258). Second, the intrinsic ATPase coupled translocase activity of the helicase domain was recently demonstrated to remove RLRs from short abundant endogenous dsRNA

(217-219). High affinity ligands, such as RNA ends harboring di- or tri-phosphates are bound by the RD and hold RIG-I on the RNA despite of ATP hydrolysis. This enables prolonged CARDs exposure sufficient for downstream signaling. Low affinity endogenous dsRNAs, however, are displaced from the helicase by ATP hydrolysis leading to increased recycling of bound RNA not able to activate IFN response. In this case the auto-repressed state of RIG-I is reestablished. Therefore, activation of RIG-I by potential endogenous RNA ligands is kinetically controlled by fast ATP turnover so that background signaling is reduced and sensitivity increased (FIG. 23).

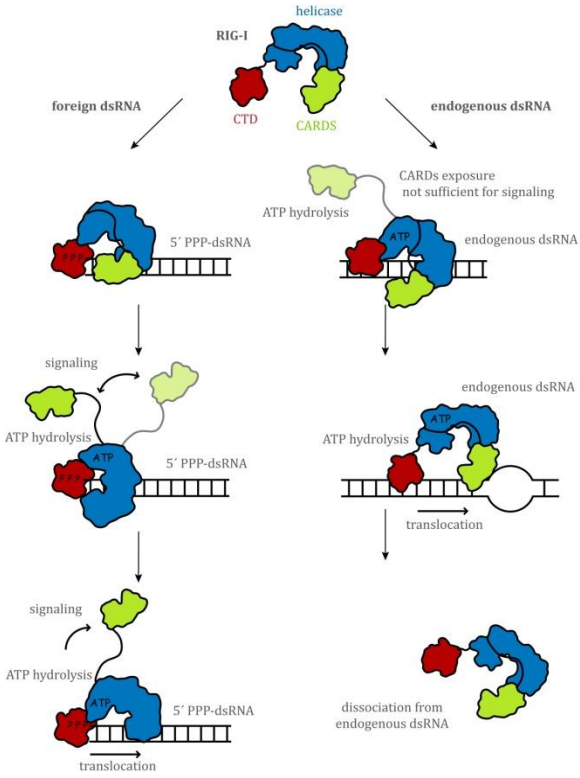


Figure 23: PROPOSED MODEL FOR ATP HYDROLYSIS OF RIG-I AS STRATEGY TO OVERCOME SELF RECOGNITION

RECOGNITION OF VIRAL RNA: RIG-I recognizes viral 5´-tri-phophorylated dsRNA via its CTD capping the dsRNA end. ATP binding to the helicase domain induces conformational changes releasing the CARDs. This high affinity ligands bound by the CTD hold RIG-I on the dsRNA despite of ATP hydrolysis. This prolonges CARDs exposure so that it is sufficient for downstream signaling. ATP hydrolysis displaces the helicase from the dsRNA and this can result in translocation along the dsRNA or rebinding by the RD at the 5'-trihophorylated dsRNA end.

RECOGNITION OF ENDOGENOUS RNA: Binding of RIG-I to low affinity endogenous dsRNA, does not lead to a sufficient exposure of the CARDs for signaling. ATP hydrolysis is enough to displace RIG-I from endogenous dsRNA, since the CTD does not provide a high affinity anchor holding RIG-I onto the dsRNA. This leads to increased recycling of bound RNA not able to active IFN response.

The proposed model for RIG-I signaling involves ATP-mediated domain compaction that relies on a steric collision between the CARDs and the RNA-bound CTD (217-219). Several evidences are provided that the CARDs can be accommodated in apo and RNA-bound states of RIG-I (201,217-219). However, when bound to RNA and ATP, conformational changes lead to the displacement of the CARDs by the RD (54,56,57). Thus, ATP binding is sufficient to force the conformational transition representing the deterministic step in RIG-I activation, whereas ATP hydrolysis is required for resetting RIG-I into the auto-repressed conformation. Maintaining of the auto-repressed state is accomplished by the pincer domain, contradicting previous results which suggested that the CTD has a repressor activity (49,217,259). Regarding RIG-I mutations related to constitutive signaling in autoimmune diseases, affinity for self-RNA coupled to slow ATP hydrolysis in RIG-I E373Q leads to a inefficient RNA displacement from the SF2 helicase domain and consequently to prolonged CARDs exposure and constitutive IFN response (217,218). Constitutively active RIG-I C268F in motif I contradicts previous mutations in motif I by mimicking a motif II mutation phenotype locking

RIG-I in a RNA-bound active state (218). In this regard, structural studies revealing the mechanistic role of C268 are required. Different mutations in the RD and as well as mutations affecting ATP binding (K270I/A) support the model that self-RNA recognition by RIG-I involves a ring-like, ATP-bound compact structure around the RNA.

RIG-I and MDA5 possess different resting states and their activation upon RNA ligand binding alongside with the effect of ATP hydrolysis is therefore as distinct as their preference for different RNA species. RIG-I exists in an auto-repressed state with a relatively flexible Hel1 domain which is maintained by the pincer, as shown in experiments in absence of the CTD (54,217). In contrast, MDA5 exists in an open conformation with a lower threshold required for activation compared to RIG-I (201). Moreover, MDA5 does not require the pincer to maintain its auto-repressed state, arguing for a different mechanism of repression in a conformation dealing with surface exposed CARDs (217). For RIG-I, the Hel2 domain was shown to couple ATP hydrolysis to translocation activity, while the Hel2i domain appeared to contact several different base pairs along the 10bp dsRNA in a series of crystal structures, suggesting a Hel2i-mediated scanning of RNA length (218,260). Further, the CARD2-Hel2i interface serves as a barrier selecting against non-blunt ended dsRNAs relying on the affinity of the dsRNA to the CTD and its ability to displace the CARDs (261). RIG-I CARDs have an inhibitory role by destabilizing RNA binding to the CTD as a well as an allosteric regulatory role. In agreement with this, removal of the CARDs increased the RNA binding affinity and ATPase activity (261). In the case of MDA5 the Hel2i accommodates the CTD of MDA5 and rigidly holds it, thereby allowing dsRNA stem binding (202). Important residues in the CARD2-Hel2i interface known from RIG-I are not conserved in MDA5, again suggesting that CARDs of MDA5 are not locked in a similar auto-repressed state. The length of the linker connecting the tandem CARDs of MDA5 is conserved and twice as long as in RIG-I. This would on one hand allow formation of oligomers composed of 8-10 CARDs compared to a CARDs tetramer for RIG-I activation (262). On the other hand a possible auto-inhibitory mechanism as with RIG-I, where a constant collision of the RD expulses the CARDs from the helicase and vice-versa, would be more intricate and probably slower. The isolated CARDs of MDA5 oligomerize in contrast to RIG-I CARDs which stay monomeric (202). A higher propensity of MDA5 CARDs to oligomerize is consistent with the requirement of more CARDs units for MDA5 signaling than RIG-I. Further, RIG-I CARDs belonging to different RIG-I dsRNA complexes have to come close together and stabilization of the CARDs is achieved by bridging of adjacent CARD2 subunits with K63-polyubiquitin chains reinforcing downstream signaling (263).

As mentioned before, MDA5 requires a lower threshold for activation, which is compensated by co-operative filament formation along one long dsRNA bringing the CARDs into close proximity. ATP binding is dispensable for MDA5 downstream signaling, however, ATP hydrolysis leads to an intrinsic kinetic instability in the filament and subsequent dissociation of MDA5 from the dsRNA (202,212). Any endogenous RNA that does not have a dsRNA structure long enough to efficiently promote nucleation of MDA5 filaments into an long-lived signaling platform, is expelled from MDA5 by its ATPase motor. Consistently, mutations in MDA5 helicase disrupting this ATPase motor result in constitutive activity as in the case of

RIG-I with MDA5 E444Q having the strongest effect (217). Interestingly, constitutive signaling of MDA5 was not abolished when combined with mutations in the RNA binding site of the CTD (217). These facts suggest that RNA length is the determining factor in MDA5 signaling activation as an alternative to the role of the CTD in RIG-I. Formation of more numerous short MDA5 filaments capable of signaling is mediated by LGP2 that facilitates nucleation of these MDA5 filaments by assisting MDA5-RNA interactions (53). LGP2 allows MDA5 filament initiation at multiple sites on a long duplex RNA over time leading to a greater number of signaling-competent MDA5 molecules. LGP2-mediated positive regulation of MDA5 signaling is dependent on ATP hydrolysis, again enforcing its important role in RLR signaling. Similar to RIG-I, LGP2 binds RNA more efficiently in the presence of ATP (264). In contrast to the other RLRs, LGP2 goes one step further by hydrolyzing ATP even in the absence of RNA. LGP2 makes use of its intrinsic basal dsRNA-independent ATP hydrolysis activity and not dsRNA-dependent hydrolysis to increase its dsRNA ligand repertoire and thus to enhance subsequent RLR activation (264). Therefore LGP2 seems to act upstream of other RLRs in order to accelerate PAMP recognition followed by improved pathogen clearance. Another role for LGP2 could be as anti-viral effector displacing viral proteins from their positions on dsRNA enabling binding of MDA5 and RIG-I to naked dsRNA. Indeed, the ATP-dependent translocase activity of RLR helicases was previously shown to remodel viral protein-dsRNA complexes through displacement of viral proteins such as NS1 (influenza A) from dsRNA (265,266). Antiviral activities of MDA5 and RIG-I can be a result of dsRNA clearing for subsequent interactions with other effector molecules such as protein kinase R (PKR) and OAS1 that then execute the antiviral activity (265). However, the effector-like functions for instance of RIG-I even extend from mere viral protein displacement to direct interference with viral replication by competing with viral polymerase P for the 5´ end of the pre-genomic RNA (267). In this case the effector function is ATP-independent since translocase activity is not required.

The results reported previously in this work, suggested that ATP hydrolysis-driven translocation results in tight packing of RIG-I along the dsRNA as seen in EM images of RIG-I filaments, instead of direct internal binding of non-translocating RIG-I which would lead to sparse and out-of-register placed individual RIG-I monomers on dsRNA (210). Moreover, enhanced signaling of RIG-I demonstrated with longer RNAs (5’ triphosphate 50bp compared to 5’ triphosphate 30bp RNA) speaks for filament formation rather than only 5´-end capping (219). However, RIG-I exclusively bound at internal sites is unable to induce downstream signaling despite of ATP hydrolysis. This was recently demonstrated by means of RIG-I binding studies on a dumbbell-like structure harboring no free ends (219). The CARDs remain in their auto-repressed conformation, suggesting that the collision force of the CTD is not enough in order to displace them. Probably in this scenario, all interactions of RIG-I with RNA are exclusively through contacts with the helicase. Availability of a duplex terminus for signaling at high concentrations of RIG-I leads to stacking of several RIG-I monomers along a RNA, thus enhancing the signaling efficiency of the end-bound RIG-I.

In spite of constant exposure and interactions to endogenous RNA stems of different nature in the cytoplasm, the receptor RIG-I remains silent due to the proofreading mechanism

enabled by ATP hydrolysis. Potential endogenous ligand RNA for RIG-I are tRNAs containing duplex RNA regions, miRNAs precursors bearing 3´overhangs, lncRNAs, specific endogenous mRNA or RNase-L cleavage products (46,261,268,269). Recently, an additional potential endogenous ligand for RLRs with long duplex RNA regions was uncovered. Several mRNA containing internal ribosomal entry sites (IRES) where identified not exclusively in viruses but also in humans (270). In this work, RIG-I was shown to bind to a ribosomal expansion segment (ES-7L) to a stretch of seven G:C/C:G bp, which approximately correspond to the footprint of dsRNA across the two helicase domains in RIG-I crystal structures (218). The accessibility of these RNA elements to cytosolic proteins combined with the high abundance of ribosomal RNA, explains why this was the predominant endogenous RNA found to bind to RIG-I E373Q (271).

3.5 NUCLEIC ACIDS SENSING AND AUTOIMMUNITY

Autoimmunity is characterized as an abnormal response of the adaptive immune system, where B cells and/ or T cells act towards endogenous antigens leading to localized tissue damage or a systemic disease. A misled adaptive immune response is often caused by aberrant PRR-triggered production of type I interferon and thus autoimmunity is tightly connected to the innate immune system. Improved understanding of innate immune signaling, especially the strategies used by nucleic acids sensing PRRs to distinguish self from non-self, would enable treatment or even prevention of specific diseases. A natural prevention mechanism of inadvertent reactions to self-nucleic acids is constituted by the rapid clearance of endogenous nucleic acids with inappropriate localization. This builds up a key checkpoint for maintaining homeostasis and control of the innate immune response. Factors antagonizing the effects through e.g. degradation of endogenous DNA in the cytosol keep the amounts of nucleic acids beneath a certain threshold and enable selective monitoring of nucleic acid sensing. Another mechanism prone to cause autoinflammatory or autoimmune syndromes is the pathogen-induced mitochondrial response, where mtDNA is released into the cytosol (117). This mechanism is beneficial for the host since it amplifies antiviral signaling and at the same time it further stimulates the thin line between foreign and self-recognition.

Alterations of the type I interferon system contributing to plasma cell differentiation and autoantibody production were found in diseases, such as systemic lupus erythematosus (SLE), Sjörgen's syndrome, diabetes mellitus type I, Aicardi-Goutières syndrome (AGS) and Singleton-Merten syndrome (SMS) (272,273). Aicardi-Goutières syndrome is a rare autoimmune disease present in neonates whose phenotype mimics an *in utero* acquired viral infection. Studies of AGS revealed at least six independent loci mutated in patients evidencing this hereditary disorder: TREX1, RNASEH2B, RNASEH2C, RNASEH2A, SAMHD1 and ADAR1 (274-277). Recently, AGS patient-related mutations were even described in the nucleic acids sensing receptor MDA5 (215,278). Nevertheless, the other mutations are as well linked to nucleic acid sensing. Mutations in the 3´-5´ exonuclease TREX1 (DNase III) are present in 25

% of the patient suffering from AGS and 1-2% of SLE patients (279). SLE is caused by a combination of multiple genetic and environmental factors that lead among others to ineffective clearance of dead cells and the respective nucleic acids. An excess supply of extracellular nucleic acids results in an aberrant immune response by nucleic acids sensors, such as cGAS, and thus leads to autoimmunity. TREX-/- mice suffer from an elevated expression of interferon induced genes (ISGs) and recent experiments demonstrated that genetic ablation of cGAS in TREX-/- MEFs and TREX-/- mice eliminated all pathological phenotypes associated with inadequate ISG induction including lethality (223,280). cGAS inhibition by small molecule drugs could provide therapeutic benefits to AGS and SLE patients with elevated cGAMP levels. Several antimalarial drugs were identified as effective dose-dependent cGAS inhibitors that disrupt dsDNA-cGAS interactions and even counteract milder manifestations of SLE (279). Seven SNPs were identified in *IFIH1* (coding MDA5) in patients with AGS and all altered single amino acids were located within the helicase domain of MDA5 (215,278). Mutations in *IFIH1* are further associated with diabetes type I, multiple sclerosis, psoriasis, selective IgA deficiency, dilated cardiomyopathy and SLE (273). Despite *IFIH1* mutations being well characterized in patients with autoimmunity, the first report correlating RIG-I with autoimmune diseases is quite recent. Mutations in DDX58 (coding RIG-I) were identified in patients with Singleton-Merten syndrome (SMS) an extremely rare multi system disorder (214). Regulation of the RLR signaling pathway is as important in the prevention of autoimmunity, as can be seen in e.g. four SNPs in MAVS associated with SLE and enhanced TRIM25 expression in peripheral blood mononuclear cells (PBMCs) derived from primary Sjögren's syndrome and in B cells from SLE patients (281,282).

Activation and regulation of the innate immune response is a proven therapeutic strategy in vaccination and its more precise understanding would improve the design of new types of effective adjuvants. Recent discoveries lead to the development of cGAMP as an endogenous potent adjuvant (114). cGAMP, as an endogenous immunostimulant, was demonstrated to boost antigen-specific T cell activation and antibody production (114). Concerning the regulation of the cGAS-STING pathway, the ecto-nucleotide pyrophosphatase/ phosphodiesterase (ENPP1) was identified as the major 2´3´-cGAMP hydrolase (283). A hydrolysis resistant variant of 2´3´-cGAMP (bisphosphothioate analog) was demonstrated to be a promising candidate agonist for vaccine adjuvant and cancer therapeutic (283). Further adjuvants connected to innate immunity are: poly I:C that enhances the adaptive immunity through activation of TLR3 and cytosolic RLRs and c-di-GMP as well as ci-di-AMP that binds to STING and was demonstrated to upregulate the expression of MHC class II and co-stimulatory molecules (284-287).

DISCUSSION

4 References

1. Wu, J., and Chen, Z. J. (2014) Innate immune sensing and signaling of cytosolic nucleic acids. *Annual review of immunology* **32**, 461-488
2. Medzhitov, R., and Janeway, C. A. (1997) Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91**, 295-298
3. Iwasaki, A., and Medzhitov, R. (2015) Control of adaptive immunity by the innate immune system. *Nature immunology* **16**, 343-353
4. Janeway, C. A. (1989) Approaching the asymptote? Evolution and revolution in immunology. in *Cold Spring Harbor symposia on quantitative biology*, Cold Spring Harbor Laboratory Press
5. Matzinger, P. (1994) Tolerance, danger, and the extended family. *Annual review of immunology* **12**, 991-1045
6. Kono, H., and Rock, K. L. (2008) How dying cells alert the immune system to danger. *Nature Reviews Immunology* **8**, 279-289
7. Lee, C. C., Avalos, A. M., and Ploegh, H. L. (2012) Accessory molecules for Toll-like receptors and their function. *Nature Reviews Immunology* **12**, 168-179
8. Janeway Jr, C. A., and Medzhitov, R. (2002) Innate immune recognition. *Annual review of immunology* **20**, 197-216
9. Chow, J., Franz, K. M., and Kagan, J. C. (2015) PRRs are watching you: Localization of innate sensing and signaling regulators. *Virology* **479**, 104-109
10. Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A. (1997) A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* **388**, 394-397
11. Anderson, K. V., Bokla, L., and Nüsslein-Volhard, C. (1985) Establishment of dorsal-ventral polarity in the Drosophila embryo: the induction of polarity by the Toll gene product. *Cell* **42**, 791-798
12. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.-M., and Hoffmann, J. A. (1996) The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell* **86**, 973-983
13. Poltorak, A., He, X., Smirnova, I., Liu, M.-Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., and Galanos, C. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085-2088
14. Kawai, T., and Akira, S. (2011) Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **34**, 637-650
15. Kawai, T., and Akira, S. (2010) The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature immunology* **11**, 373-384
16. Alexopoulou, L., Holt, A. C., Medzhitov, R., and Flavell, R. A. (2001) Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3. *Nature* **413**, 732-738
17. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S., and e Sousa, C. R. (2004) Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529-1531
18. Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S. (2004) Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526-1529

19. Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., and Takeda, K. (2000) A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740-745
20. Li, X.-D., and Chen, Z. J. (2012) Sequence specific detection of bacterial 23S ribosomal RNA by TLR13. *Elife* **1**, e00102
21. Latz, E., Verma, A., Visintin, A., Gong, M., Sirois, C. M., Klein, D. C., Monks, B. G., McKnight, C. J., Lamphier, M. S., and Duprex, W. P. (2007) Ligand-induced conformational changes allosterically activate Toll-like receptor 9. *Nature immunology* **8**, 772-779
22. Takeuchi, O., and Akira, S. (2010) Pattern recognition receptors and inflammation. *Cell* **140**, 805-820
23. Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006) Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**, 803-814
24. Shaw, M. H., Reimer, T., Kim, Y.-G., and Nuñez, G. (2008) NOD-like receptors (NLRs): bona fide intracellular microbial sensors. *Current opinion in immunology* **20**, 377-382
25. Franchi, L., Eigenbrod, T., Muñoz-Planillo, R., and Nuñez, G. (2009) The inflammasome: a caspase-1-activation platform that regulates immune responses and disease pathogenesis. *Nature immunology* **10**, 241-247
26. Zhou, R., Yazdi, A. S., Menu, P., and Tschopp, J. (2011) A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469**, 221-225
27. Zhang, L., Chen, S., Ruan, J., Wu, J., Tong, A. B., Yin, Q., Li, Y., David, L., Lu, A., and Wang, W. L. (2015) Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* **350**, 404-409
28. Hu, Z., Zhou, Q., Zhang, C., Fan, S., Cheng, W., Zhao, Y., Shao, F., Wang, H.-W., Sui, S.-F., and Chai, J. (2015) Structural and biochemical basis for induced self-propagation of NLRC4. *Science* **350**, 399-404
29. Hong, M., Yoon, S.-i., and Wilson, I. A. (2012) Structure and functional characterization of the RNA-binding element of the NLRX1 innate immune modulator. *Immunity* **36**, 337-347
30. Diebold, S. S., Montoya, M., Unger, H., Alexopoulou, L., Roy, P., Haswell, L. E., Al-Shamkhani, A., Flavell, R., Borrow, P., and e Sousa, C. R. (2003) Viral infection switches non-plasmacytoid dendritic cells into high interferon producers. *Nature* **424**, 324-328
31. Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature immunology* **5**, 730-737
32. Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y.-M., Gale, M., and Akira, S. (2005) Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *The Journal of Immunology* **175**, 2851-2858
33. Gitlin, L., Barchet, W., Gilfillan, S., Cella, M., Beutler, B., Flavell, R. A., Diamond, M. S., and Colonna, M. (2006) Essential role of mda-5 in type I IFN responses to polyriboinosinic: polyribocytidylic acid and encephalomyocarditis picornavirus. *Proceedings of the National Academy of Sciences* **103**, 8459-8464
34. Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., and Ishii, K. J. (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101-105
35. Fredericksen, B. L., Keller, B. C., Fornek, J., Katze, M. G., and Gale, M. (2008) Establishment and maintenance of the innate antiviral response to West Nile Virus

involves both RIG-I and MDA5 signaling through IPS-1. *Journal of virology* **82**, 609-616

36. Loo, Y.-M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira, S., Gill, M. A., García-Sastre, A., and Katze, M. G. (2008) Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *Journal of virology* **82**, 335-345
37. Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T. S., Fujita, T., and Akira, S. (2008) Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid–inducible gene-I and melanoma differentiation–associated gene 5. *The Journal of experimental medicine* **205**, 1601-1610
38. Hornung, V., Ellegast, J., Kim, S., Brzózka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.-K., and Schlee, M. (2006) 5'-Triphosphate RNA is the ligand for RIG-I. *Science* **314**, 994-997
39. Pichlmair, A., Schulz, O., Tan, C. P., Näslund, T. I., Liljeström, P., Weber, F., and e Sousa, C. R. (2006) RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**, 997-1001
40. Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A. J., Zillinger, T., Goldeck, M., Schuberth, C., Van der Veen, A. G., Fujimura, T., and Rehwinkel, J. (2014) Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5 [prime]-diphosphates. *Nature* **514**, 372-375
41. Devarkar, S. C., Wang, C., Miller, M. T., Ramanathan, A., Jiang, F., Khan, A. G., Patel, S. S., and Marcotrigiano, J. (2016) Structural basis for m7G recognition and 2′-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proceedings of the National Academy of Sciences*, 201515152
42. Schuberth-Wagner, C., Ludwig, J., Bruder, A. K., Herzner, A.-M., Zillinger, T., Goldeck, M., Schmidt, T., Schmid-Burgk, J. L., Kerber, R., and Wolter, S. (2015) A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N 1-2′ O-Methylated Self RNA. *Immunity* **43**, 41-51
43. Schlee, M., Roth, A., Hornung, V., Hagmann, C. A., Wimmenauer, V., Barchet, W., Coch, C., Janke, M., Mihailovic, A., and Wardle, G. (2009) Recognition of 5′ triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunity* **31**, 25-34
44. Schmidt, A., Schwerd, T., Hamm, W., Hellmuth, J. C., Cui, S., Wenzel, M., Hoffmann, F. S., Michallet, M.-C., Besch, R., and Hopfner, K.-P. (2009) 5′-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I. *Proceedings of the National Academy of Sciences* **106**, 12067-12072
45. Liu, G., Park, H.-S., Pyo, H.-M., Liu, Q., and Zhou, Y. (2015) Influenza A Virus Panhandle Structure Is Directly Involved in RIG-I Activation and Interferon Induction. *Journal of virology* **89**, 6067-6079
46. Malathi, K., Dong, B., Gale, M., and Silverman, R. H. (2007) Small self-RNA generated by RNase L amplifies antiviral innate immunity. *Nature* **448**, 816-819
47. Pichlmair, A., Schulz, O., Tan, C.-P., Rehwinkel, J., Kato, H., Takeuchi, O., Akira, S., Way, M., Schiavo, G., and e Sousa, C. R. (2009) Activation of MDA5 requires higher-order RNA structures generated during virus infection. *Journal of virology* **83**, 10761-10769
48. Bamming, D., and Horvath, C. M. (2009) Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *Journal of Biological Chemistry* **284**, 9700-9712
49. Saito, T., Hirai, R., Loo, Y.-M., Owen, D., Johnson, C. L., Sinha, S. C., Akira, S., Fujita, T., and Gale, M. (2007) Regulation of innate antiviral defenses through a shared

repressor domain in RIG-I and LGP2. *Proceedings of the National Academy of Sciences* **104**, 582-587

50. Pippig, D. A., Hellmuth, J. C., Cui, S., Kirchhofer, A., Lammens, K., Lammens, A., Schmidt, A., Rothenfusser, S., and Hopfner, K.-P. (2009) The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA. *Nucleic acids research* **37**, 2014-2025
51. Venkataraman, T., Valdes, M., Elsby, R., Kakuta, S., Caceres, G., Saijo, S., Iwakura, Y., and Barber, G. N. (2007) Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. *The Journal of Immunology* **178**, 6444-6455
52. Satoh, T., Kato, H., Kumagai, Y., Yoneyama, M., Sato, S., Matsushita, K., Tsujimura, T., Fujita, T., Akira, S., and Takeuchi, O. (2010) LGP2 is a positive regulator of RIG-I–and MDA5-mediated antiviral responses. *Proceedings of the National Academy of Sciences* **107**, 1512-1517
53. Bruns, A. M., Leser, G. P., Lamb, R. A., and Horvath, C. M. (2014) The innate immune sensor LGP2 activates antiviral signaling by regulating MDA5-RNA interaction and filament assembly. *Molecular cell* **55**, 771-781
54. Kowalinski, E., Lunardi, T., McCarthy, A. A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D., and Cusack, S. (2011) Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**, 423-435
55. Leung, D. W., and Amarasinghe, G. K. (2012) Structural insights into RNA recognition and activation of RIG-I-like receptors. *Current opinion in structural biology* **22**, 297-303
56. Jiang, F., Ramanathan, A., Miller, M. T., Tang, G.-Q., Gale, M., Patel, S. S., and Marcotrigiano, J. (2011) Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* **479**, 423-427
57. Luo, D., Ding, S. C., Vela, A., Kohlway, A., Lindenbach, B. D., and Pyle, A. M. (2011) Structural insights into RNA recognition by RIG-I. *Cell* **147**, 409-422
58. Civril, F., Bennett, M., Moldt, M., Deimling, T., Witte, G., Schiesser, S., Carell, T., and Hopfner, K. P. (2011) The RIG-I ATPase domain structure reveals insights into ATP-dependent antiviral signalling. *EMBO reports* **12**, 1127-1134
59. Gack, M. U., Shin, Y. C., Joo, C.-H., Urano, T., Liang, C., Sun, L., Takeuchi, O., Akira, S., Chen, Z., and Inoue, S. (2007) TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446**, 916-920
60. Oshiumi, H., Matsumoto, M., Hatakeyama, S., and Seya, T. (2009) Riplet/RNF135, a RING finger protein, ubiquitinates RIG-I to promote interferon-β induction during the early phase of viral infection. *Journal of Biological Chemistry* **284**, 807-817
61. Jiang, X., Kinch, L. N., Brautigam, C. A., Chen, X., Du, F., Grishin, N. V., and Chen, Z. J. (2012) Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity* **36**, 959-973
62. Seth, R. B., Sun, L., Ea, C.-K., and Chen, Z. J. (2005) Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-κB and IRF3. *Cell* **122**, 669-682
63. Kawai, T., Takahashi, K., Sato, S., Coban, C., Kumar, H., Kato, H., Ishii, K. J., Takeuchi, O., and Akira, S. (2005) IPS-1, an adaptor triggering RIG-I-and Mda5-mediated type I interferon induction. *Nature immunology* **6**, 981-988
64. Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R., and Tschopp, J. (2005) Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167-1172

65. Xu, L.-G., Wang, Y.-Y., Han, K.-J., Li, L.-Y., Zhai, Z., and Shu, H.-B. (2005) VISA is an adapter protein required for virus-triggered IFN-β signaling. *Molecular cell* **19**, 727-740
66. Hou, F., Sun, L., Zheng, H., Skaug, B., Jiang, Q.-X., and Chen, Z. J. (2011) MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* **146**, 448-461
67. Wu, B., Peisley, A., Tetrault, D., Li, Z., Egelman, E. H., Magor, K. E., Walz, T., Penczek, P. A., and Hur, S. (2014) Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Molecular cell* **55**, 511-523
68. Xu, H., He, X., Zheng, H., Huang, L. J., Hou, F., Yu, Z., de la Cruz, M. J., Borkowski, B., Zhang, X., and Chen, Z. J. (2014) Structural basis for the prion-like MAVS filaments in antiviral innate immunity. *Elife* **3**, e01489
69. Zeng, W., Sun, L., Jiang, X., Chen, X., Hou, F., Adhikari, A., Xu, M., and Chen, Z. J. (2010) Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. *Cell* **141**, 315-330
70. Liu, S., Chen, J., Cai, X., Wu, J., Chen, X., Wu, Y.-T., Sun, L., and Chen, Z. J. (2013) MAVS recruits multiple ubiquitin E3 ligases to activate antiviral signaling cascades. *Elife* **2**, e00785
71. Saha, S. K., Pietras, E. M., He, J. Q., Kang, J. R., Liu, S. Y., Oganesyan, G., Shahangian, A., Zarnegar, B., Shiba, T. L., and Wang, Y. (2006) Regulation of antiviral responses by a direct and specific interaction between TRAF3 and Cardif. *The EMBO journal* **25**, 3257-3263
72. Tang, E. D., and Wang, C.-Y. (2010) TRAF5 is a downstream target of MAVS in antiviral innate immune signaling. *PLoS One* **5**, e9172
73. Paz, S., Vilasco, M., Werden, S. J., Arguello, M., Joseph-Pillai, D., Zhao, T., Nguyen, T. L.-A., Sun, Q., Meurs, E. F., and Lin, R. (2011) A functional C-terminal TRAF3-binding site in MAVS participates in positive and negative regulation of the IFN antiviral response. *Cell research* **21**, 895-910
74. Wu, C.-J., Conze, D. B., Li, T., Srinivasula, S. M., and Ashwell, J. D. (2006) Sensing of Lys 63-linked polyubiquitination by NEMO is a key event in NF-κB activation. *Nature cell biology* **8**, 398-406
75. Ea, C.-K., Deng, L., Xia, Z.-P., Pineda, G., and Chen, Z. J. (2006) Activation of IKK by TNFα requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. *Molecular cell* **22**, 245-257
76. Chebath, J., Benech, P., Revel, M., and Vigneron, M. (1987) Constitutive expression of (2′–5′) oligo A synthetase confers resistance to picornavirus infection.
77. Benech, P., Vigneron, M., Peretz, D., Revel, M., and Chebath, J. (1987) Interferon-responsive regulatory elements in the promoter of the human 2', 5'-oligo (A) synthetase gene. *Molecular and cellular biology* **7**, 4498-4504
78. Minks, M. A., West, D. K., Benvin, S., and Baglioni, C. (1979) Structural requirements of double-stranded RNA for the activation of 2', 5'-oligo (A) polymerase and protein kinase of interferon-treated HeLa cells. *Journal of Biological Chemistry* **254**, 10180-10183
79. Hovanessian, A. G., Brown, R. E., and Kerr, I. M. (1977) Synthesis of low molecular weight inhibitor of protein synthesis with enzyme from interferon-treated cells.
80. Roberts, W. K., Hovanessian, A., Brown, R. E., Clemens, M. J., and Kerr, I. M. (1976) Interferon-mediated protein kinase and low-molecular-weight inhibitor of protein synthesis.
81. Kerr, I. M., Brown, R. E., and Hovanessian, A. G. (1977) Nature of inhibitor of cell-free protein synthesis formed in response to interferon and double-stranded RNA.

82. Baglioni, C., Minks, M. A., and Maroney, P. A. (1978) Interferon action may be mediated by activation of a nuclease by pppA2′ p5′ A2′ p5′ A.
83. Zilberstein, A., Kimchi, A., Schmidt, A., and Revel, M. (1978) Isolation of two interferon-induced translational inhibitors: a protein kinase and an oligo-isoadenylate synthetase. *Proceedings of the National Academy of Sciences* **75**, 4734-4738
84. Malathi, K., Saito, T., Crochet, N., Barton, D. J., Gale, M., and Silverman, R. H. (2010) RNase L releases a small RNA from HCV RNA that refolds into a potent PAMP. *Rna* **16**, 2108-2119
85. Rehwinkel, J., Tan, C. P., Goubau, D., Schulz, O., Pichlmair, A., Bier, K., Robb, N., Vreede, F., Barclay, W., and Fodor, E. (2010) RIG-I detects viral genomic RNA during negative-strand RNA virus infection. *Cell* **140**, 397-408
86. Chebath, J., Benech, P., Hovanessian, A., Galabru, J., and Revel, M. (1987) Four different forms of interferon-induced 2', 5'-oligo (A) synthetase identified by immunoblotting in human cells. *Journal of Biological Chemistry* **262**, 3852-3857
87. Ghosh, A., Sarkar, S. N., Guo, W., Bandyopadhyay, S., and Sen, G. C. (1997) Enzymatic activity of 2′–5′-oligoadenylate synthetase is impaired by specific mutations that affect oligomerization of the protein. *Journal of Biological Chemistry* **272**, 33220-33226
88. Rebouillat, D., Marié, I., and Hovanessian, A. G. (1998) Molecular cloning and characterization of two related and interferon-induced 56-kDa and 30-kDa proteins highly similar to 2′-5′ oligoadenylate synthetase. *European Journal of Biochemistry* **257**, 319-330
89. Marques, J., Anwar, J., Eskildsen-Larsen, S., Rebouillat, D., Paludan, S. R., Sen, G., Williams, B. R., and Hartmann, R. (2008) The p59 oligoadenylate synthetase-like protein possesses antiviral activity that requires the C-terminal ubiquitin-like domain. *Journal of General Virology* **89**, 2767-2772
90. Liu, S.-Y., Sanchez, D. J., Aliyari, R., Lu, S., and Cheng, G. (2012) Systematic identification of type I and type II interferon-induced antiviral factors. *Proceedings of the National Academy of Sciences* **109**, 4239-4244
91. Guo, X., Li, X., Xu, Y., Sun, T., Yang, G., Wu, Z., and Li, E. (2012) Identification of OASL d, a splice variant of human OASL, with antiviral activity. *The international journal of biochemistry & cell biology* **44**, 1133-1138
92. Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., and Rice, C. M. (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481-485
93. Ibsen, M. S., Gad, H. H., Andersen, L. L., Hornung, V., Julkunen, I., Sarkar, S. N., and Hartmann, R. (2015) Structural and functional analysis reveals that human OASL binds dsRNA to enhance RIG-I signaling. *Nucleic acids research*, gkv389
94. Zhu, J., Zhang, Y., Ghosh, A., Cuevas, R. A., Forero, A., Dhar, J., Ibsen, M. S., Schmid-Burgk, J. L., Schmidt, T., and Ganapathiraju, M. K. (2014) Antiviral activity of human OASL protein is mediated by enhancing signaling of the RIG-I RNA sensor. *Immunity* **40**, 936-948
95. O'Neill, L. (2013) Sensing the dark side of DNA. *Science* **339**, 763-764
96. Ishikawa, H., and Barber, G. N. (2008) STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455**, 674-678
97. Sauer, J.-D., Sotelo-Troha, K., von Moltke, J., Monroe, K. M., Rae, C. S., Brubaker, S. W., Hyodo, M., Hayakawa, Y., Woodward, J. J., and Portnoy, D. A. (2011) The N-ethyl-N-nitrosourea-induced Goldenticket mouse mutant reveals an essential function of Sting

in the in vivo interferon response to Listeria monocytogenes and cyclic dinucleotides. *Infection and immunity* **79**, 688-694

98. Jin, L., Hill, K. K., Filak, H., Mogan, J., Knowles, H., Zhang, B., Perraud, A.-L., Cambier, J. C., and Lenz, L. L. (2011) MPYS is required for IFN response factor 3 activation and type I IFN production in the response of cultured phagocytes to bacterial second messengers cyclic-di-AMP and cyclic-di-GMP. *The Journal of Immunology* **187**, 2595-2601
99. Burdette, D. L., Monroe, K. M., Sotelo-Troha, K., Iwig, J. S., Eckert, B., Hyodo, M., Hayakawa, Y., and Vance, R. E. (2011) STING is a direct innate immune sensor of cyclic di-GMP. *Nature* **478**, 515-518
100. Yin, Q., Tian, Y., Kabaleeswaran, V., Jiang, X., Tu, D., Eck, M. J., Chen, Z. J., and Wu, H. (2012) Cyclic di-GMP sensing via the innate immune signaling protein STING. *Molecular cell* **46**, 735-745
101. Tanaka, Y., and Chen, Z. J. (2012) STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Science signaling* **5**, ra20
102. Burdette, D. L., and Vance, R. E. (2013) STING and the innate immune response to nucleic acids in the cytosol. *Nature immunology* **14**, 19-26
103. Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z. J. (2013) Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786-791
104. Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., and Chen, Z. J. (2013) Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826-830
105. Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Röhl, I., Hopfner, K.-P., Ludwig, J., and Hornung, V. (2013) cGAS produces a 2 [prime]-5 [prime]-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380-384
106. Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B. L., Zillinger, T., Serganov, A. A., Liu, Y., Jones, R. A., and Hartmann, G. (2013) Cyclic [G (2′, 5′) pA (3′, 5′) p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153**, 1094-1107
107. Diner, E. J., Burdette, D. L., Wilson, S. C., Monroe, K. M., Kellenberger, C. A., Hyodo, M., Hayakawa, Y., Hammond, M. C., and Vance, R. E. (2013) The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell reports* **3**, 1355-1361
108. Zhang, X., Shi, H., Wu, J., Zhang, X., Sun, L., Chen, C., and Chen, Z. J. (2013) Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Molecular cell* **51**, 226-235
109. Li, X., Shu, C., Yi, G., Chaton, C. T., Shelton, C. L., Diao, J., Zuo, X., Kao, C. C., Herr, A. B., and Li, P. (2013) Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019-1031
110. Gao, P., Ascano, M., Zillinger, T., Wang, W., Dai, P., Serganov, A. A., Gaffney, B. L., Shuman, S., Jones, R. A., and Deng, L. (2013) Structure-function analysis of STING activation by c [G (2′, 5′) pA (3′, 5′) p] and targeting by antiviral DMXAA. *Cell* **154**, 748-762
111. Ablasser, A., Schmid-Burgk, J. L., Hemmerling, I., Horvath, G. L., Schmidt, T., Latz, E., and Hornung, V. (2013) Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP. *Nature* **503**, 530-534
112. Gentili, M., Kowal, J., Tkach, M., Satoh, T., Lahaye, X., Conrad, C., Boyron, M., Lombard, B., Durand, S., and Kroemer, G. (2015) Transmission of innate immune signaling by packaging of cGAMP in viral particles. *Science* **349**, 1232-1236

113. Bridgeman, A., Maelfait, J., Davenne, T., Partridge, T., Peng, Y., Mayer, A., Dong, T., Kaever, V., Borrow, P., and Rehwinkel, J. (2015) Viruses transfer the antiviral second messenger cGAMP between cells. *Science* **349**, 1228-1232
114. Li, X.-D., Wu, J., Gao, D., Wang, H., Sun, L., and Chen, Z. J. (2013) Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* **341**, 1390-1394
115. Gao, D., Wu, J., Wu, Y.-T., Du, F., Aroh, C., Yan, N., Sun, L., and Chen, Z. J. (2013) Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341**, 903-906
116. Schoggins, J. W., MacDuff, D. A., Imanaka, N., Gainey, M. D., Shrestha, B., Eitson, J. L., Mar, K. B., Richardson, R. B., Ratushny, A. V., and Litvak, V. (2014) Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* **505**, 691-695
117. West, A. P., Khoury-Hanold, W., Staron, M., Tal, M. C., Pineda, C. M., Lang, S. M., Bestwick, M., Duguay, B. A., Raimundo, N., and MacDuff, D. A. (2015) Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520**, 553-557
118. Rongvaux, A., Jackson, R., Harman, C. C., Li, T., West, A. P., de Zoete, M. R., Wu, Y., Yordy, B., Lakhani, S. A., and Kuan, C.-Y. (2014) Apoptotic caspases prevent the induction of type I interferons by mitochondrial DNA. *Cell* **159**, 1563-1577
119. Lau, L., Gray, E. E., Brunette, R. L., and Stetson, D. B. (2015) DNA tumor virus oncogenes antagonize the cGAS-STING DNA-sensing pathway. *Science* **350**, 568-571
120. Takaoka, A., Wang, Z., Choi, M. K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., and Honda, K. (2007) DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501-505
121. Wang, Z., Choi, M. K., Ban, T., Yanai, H., Negishi, H., Lu, Y., Tamura, T., Takaoka, A., Nishikura, K., and Taniguchi, T. (2008) Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules. *Proceedings of the National Academy of Sciences* **105**, 5477-5482
122. Ishii, K. J., Kawagoe, T., Koyama, S., Matsui, K., Kumar, H., Kawai, T., Uematsu, S., Takeuchi, O., Takeshita, F., and Coban, C. (2008) TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* **451**, 725-729
123. Unterholzner, L., Keating, S. E., Baran, M., Horan, K. A., Jensen, S. B., Sharma, S., Sirois, C. M., Jin, T., Latz, E., and Xiao, T. S. (2010) IFI16 is an innate immune sensor for intracellular DNA. *Nature immunology* **11**, 997-1004
124. Roberts, T. L., Idris, A., Dunn, J. A., Kelly, G. M., Burnton, C. M., Hodgson, S., Hardy, L. L., Garceau, V., Sweet, M. J., and Ross, I. L. (2009) HIN-200 proteins regulate caspase activation in response to foreign cytoplasmic DNA. *Science* **323**, 1057-1060
125. Fernandes-Alnemri, T., Yu, J.-W., Datta, P., Wu, J., and Alnemri, E. S. (2009) AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* **458**, 509-513
126. Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D. R., Latz, E., and Fitzgerald, K. A. (2009) AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **458**, 514-518
127. Bürckstümmer, T., Baumann, C., Blüml, S., Dixit, E., Dürnberger, G., Jahn, H., Planyavsky, M., Bilban, M., Colinge, J., and Bennett, K. L. (2009) An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nature immunology* **10**, 266-272
128. Morrone, S. R., Wang, T., Constantoulakis, L. M., Hooy, R. M., Delannoy, M. J., and Sohn, J. (2014) Cooperative assembly of IFI16 filaments on dsDNA provides insights into host defense strategy. *Proceedings of the National Academy of Sciences* **111**, E62-E71

129. Lu, A., Kabaleeswaran, V., Fu, T., Magupalli, V. G., and Wu, H. (2014) Crystal structure of the F27G AIM2 PYD mutant and similarities of its self-association to DED/DED interactions. *Journal of molecular biology* **426**, 1420-1427
130. Zhang, Z., Yuan, B., Bao, M., Lu, N., Kim, T., and Liu, Y.-J. (2011) The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nature immunology* **12**, 959-965
131. Parvatiyar, K., Zhang, Z., Teles, R. M., Ouyang, S., Jiang, Y., Iyer, S. S., Zaver, S. A., Schenk, M., Zeng, S., and Zhong, W. (2012) The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. *Nature immunology* **13**, 1155-1161
132. Ferguson, B. J., Mansur, D. S., Peters, N. E., Ren, H., and Smith, G. L. (2012) DNA-PK is a DNA sensor for IRF-3-dependent innate immunity. *elife* **1**, e00047
133. Imamichi, T., Zhang, X., Zhou, M., Lempicki, R., Baseler, M., Veenstra, T., Young, H., and Lane, H. C. (2011) Ku70 is a novel cytosolic DNA sensor that induces a Type-III rather than Type-I IFN via activation of IRF-1 and IRF-7. *The Journal of Immunology* **186**, 116.111
134. Kondo, T., Kobayashi, J., Saitoh, T., Maruyama, K., Ishii, K. J., Barber, G. N., Komatsu, K., Akira, S., and Kawai, T. (2013) DNA damage sensor MRE11 recognizes cytosolic double-stranded DNA and induces type I interferon by regulating STING trafficking. *Proceedings of the National Academy of Sciences* **110**, 2969-2974
135. Civril, F., Deimling, T., de Oliveira Mann, C. C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.-P. (2013) Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332-337
136. Kranzusch, P. J., Lee, A. S.-Y., Berger, J. M., and Doudna, J. A. (2013) Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell reports* **3**, 1362-1368
137. Kato, K., Ishii, R., Goto, E., Ishitani, R., Tokunaga, F., and Nureki, O. (2013) Structural and functional analyses of DNA-sensing and immune activation by human cGAS. *PloS one* **8**, e76983
138. Donovan, J., Dufner, M., and Korennykh, A. (2013) Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proceedings of the National Academy of Sciences* **110**, 1652-1657
139. Zhang, X., Wu, J., Du, F., Xu, H., Sun, L., Chen, Z., Brautigam, C. A., Zhang, X., and Chen, Z. J. (2014) The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell reports* **6**, 421-430
140. Steitz, T., Smerdon, S., Jager, J., and Joyce, C. (1994) A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* **266**, 2022-2025
141. Steitz, T. A. (1999) DNA polymerases: structural diversity and common mechanisms. *Journal of Biological Chemistry* **274**, 17395-17398
142. Witte, G., Hartung, S., Büttner, K., and Hopfner, K.-P. (2008) Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Molecular cell* **30**, 167-178
143. Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., and Schirmer, T. (2004) Structural basis of activity and allosteric control of diguanylate cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17084-17089
144. Mozzi, A., Pontremoli, C., Forni, D., Clerici, M., Pozzoli, U., Bresolin, N., Cagliani, R., and Sironi, M. (2015) OASes and STING: adaptive evolution in concert. *Genome biology and evolution* **7**, 1016-1032

145. Hornung, V., Hartmann, R., Ablasser, A., and Hopfner, K.-P. (2014) OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nature Reviews Immunology* **14**, 521-528
146. Lohöfener, J., Steinke, N., Kay-Fedorov, P., Baruch, P., Nikulin, A., Tishchenko, S., Manstein, D. J., and Fedorov, R. (2015) The Activation Mechanism of 2′-5′-Oligoadenylate Synthetase Gives New Insights Into OAS/cGAS Triggers of Innate Immunity. *Structure* **23**, 851-862
147. Hancks, D. C., Hartley, M. K., Hagan, C., Clark, N. L., and Elde, N. C. (2015) Overlapping Patterns of Rapid Evolution in the Nucleic Acid Sensors cGAS and OAS1 Suggest a Common Mechanism of Pathogen Antagonism and Escape. *PLoS Genet* **11**, e1005203
148. Holm, L., and Sander, C. (1995) DNA polymerase β belongs to an ancient nucleotidyltransferase superfamily. *Trends in biochemical sciences* **20**, 345-347
149. Aravind, L., and Koonin, E. V. (1999) DNA polymerase β-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic acids research* **27**, 1609-1618
150. Kuchta, K., Knizewski, L., Wyrwicz, L. S., Rychlewski, L., and Ginalski, K. (2009) Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic acids research* **37**, 7701-7714
151. Ho, S. H., So, G. M., and Chow, K. L. (2001) Postembryonic expression of Caenorhabditis elegans mab-21 and its requirement in sensory ray differentiation. *Developmental Dynamics* **221**, 422-430
152. Kao, C., Chiang, Y., Chen, P., Lin, K., Hwang, P., Yang-Yen, H., and Yen, J. J. (2008) CBAP interacts with the un-liganded common β-subunit of the GM-CSF/IL-3/IL-5 receptor and induces apoptosis via mitochondrial dysfunction. *Oncogene* **27**, 1397-1403
153. Chiang, Y.-J., Ho, K.-C., Sun, C.-T., Chiu, J.-J., Lee, F.-J., Liao, F., Yang-Yen, H.-F., and Yen, J. J.-Y. (2013) CBAP functions as a novel component in chemokine-induced ZAP70-mediated T-cell adhesion and migration. *PloS one* **8**, e61761
154. Ho, K., Chiang, Y., Lai, A. C., Liao, N., Chang, Y., Yang-Yen, H., and Yen, J. J. (2014) CBAP promotes thymocyte negative selection by facilitating T-cell receptor proximal signaling. *Cell death & disease* **5**, e1518
155. van Rossum, D. B., Patterson, R. L., Cheung, K.-H., Barrow, R. K., Syrovatkina, V., Gessell, G. S., Burkholder, S. G., Watkins, D. N., Foskett, J. K., and Snyder, S. H. (2006) DANGER, a novel regulatory protein of inositol 1, 4, 5-trisphosphate-receptor activity. *Journal of Biological Chemistry* **281**, 37111-37116
156. Kang, B. N., Ahmad, A. S., Saleem, S., Patterson, R. L., Hester, L., Doré, S., and Snyder, S. H. (2010) Death-associated protein kinase-mediated cell death modulated by interaction with DANGER. *The Journal of Neuroscience* **30**, 93-98
157. Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R., and Wiemann, S. (2000) Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO reports* **1**, 287-292
158. Palmer, C. S., Osellame, L. D., Laine, D., Koutsopoulos, O. S., Frazier, A. E., and Ryan, M. T. (2011) MiD49 and MiD51, new components of the mitochondrial fission machinery. *EMBO reports* **12**, 565-573
159. Losón, O. C., Song, Z., Chen, H., and Chan, D. C. (2013) Fis1, Mff, MiD49, and MiD51 mediate Drp1 recruitment in mitochondrial fission. *Molecular biology of the cell* **24**, 659-667
160. Palmer, C. S., Elgass, K. D., Parton, R. G., Osellame, L. D., Stojanovski, D., and Ryan, M. T. (2013) Adaptor proteins MiD49 and MiD51 can act independently of Mff and Fis1 in

Drp1 recruitment and are specific for mitochondrial fission. *Journal of Biological Chemistry* **288**, 27584-27593

161. Richter, V., Palmer, C. S., Osellame, L. D., Singh, A. P., Elgass, K., Stroud, D. A., Sesaki, H., Kvansakul, M., and Ryan, M. T. (2014) Structural and functional analysis of MiD51, a dynamin receptor required for mitochondrial fission. *The Journal of cell biology* **204**, 477-486
162. Losón, O. C., Liu, R., Rome, M. E., Meng, S., Kaiser, J. T., Shan, S.-o., and Chan, D. C. (2014) The mitochondrial fission receptor MiD51 requires ADP as a cofactor. *Structure* **22**, 367-377
163. Losón, O. C., Meng, S., Ngo, H., Liu, R., Kaiser, J. T., and Chan, D. C. (2015) Crystal structure and functional analysis of MiD49, a receptor for the mitochondrial fission protein Drp1. *Protein Science* **24**, 386-394
164. Chow, K. L., Hall, D. H., and Emmons, S. W. (1995) The mab-21 gene of Caenorhabditis elegans encodes a novel protein required for choice of alternate cell fates. *Development* **121**, 3615-3626
165. Baird, S. E., Fitch, D., Kassem, I., and Emmons, S. W. (1991) Pattern formation in the nematode epidermis: determination of the arrangement of peripheral sense organs in the C. elegans male tail. *Development* **113**, 515-526
166. Morita, K., Chow, K. L., and Ueno, N. (1999) Regulation of body length and male tail ray pattern formation of Caenorhabditis elegans by a member of TGF-beta family. *Development* **126**, 1337-1347
167. Suzuki, Y., Yandell, M. D., Roy, P. J., Krishna, S., Savage-Dunn, C., Ross, R. M., Padgett, R. W., and Wood, W. B. (1999) A BMP homolog acts as a dose-dependent regulator of body size and male tail patterning in Caenorhabditis elegans. *Development* **126**, 241-250
168. Margolis, R. L., Stine, O. C., McInnis, M. G., Ranen, N. G., Rubinsztein, D. C., Leggo, J., Brando, L. V. J., Kidwai, A. S., Loev, S. J., and Breschel, T. S. (1996) cDNA cloning of a human homologue of the Caenorhabditis elegans cell fate-determining gene mab-21: expression, chromosomal localization and analysis of a highly polymorphic (CAG) n trinucleotide repeat. *Human molecular genetics* **5**, 607-616
169. Wong, R., Wong, H., and Chow, K. (1999) Genomic cloning and chromosomal localization of the mouse Mab21l2 locus. *Cytogenetic and Genome Research* **86**, 21-24
170. Mariani, M., Baldessari, D., Francisconi, S., Viggiano, L., Rocchi, M., Zappavigna, V., Malgaretti, N., and Consalez, G. G. (1999) Two murine and human homologs of mab-21, a cell fate determination gene involved in Caenorhabditis elegans neural development. *Human molecular genetics* **8**, 2397-2406
171. Tsang, W., Shek, K., Lee, T., and Chow, K. (2009) An evolutionarily conserved nested gene pair—Mab21 and Lrba/Nbea in metazoan. *Genomics* **94**, 177-187
172. Mariani, M., Corradi, A., Baldessari, D., Malgaretti, N., Pozzoli, O., Fesce, R., Martinez, S., Boncinelli, E., and Consalez, G. (1998) Mab21, the mouse homolog of a C. elegans cell-fate specification gene, participates in cerebellar, midbrain and eye development. *Mechanisms of development* **79**, 131-135
173. Wong, R. L. Y., Chan, K. K. L., and Chow, K. L. (1999) Developmental expression of Mab21l2 during mouse embryogenesis. *Mechanisms of development* **87**, 185-188
174. Lau, G. T., Wong, O. G., Chan, P. M., Kok, K.-H., Wong, R. L., Chin, K.-T., Lin, M. C., Kung, H.-F., and Chow, K. L. (2001) Embryonic XMab21l2 expression is required for gastrulation and subsequent neural development. *Biochemical and biophysical research communications* **280**, 1378-1384

175. Kudoh, T., and Dawid, I. B. (2001) Zebrafish mab21l2 is specifically expressed in the presumptive eye and tectum from early somitogenesis onwards. *Mechanisms of development* **109**, 95-98
176. Wong, Y.-M., and Chow, K. L. (2002) Expression of zebrafish mab21 genes marks the differentiating eye, midbrain and neural tube. *Mechanisms of development* **113**, 149-152
177. Cederlund, M. L., Vendrell, V., Morrissey, M. E., Yin, J., Gaora, P. Ó., Smyth, V. A., Higgins, D. G., and Kennedy, B. N. (2011) mab21l2 transgenics reveal novel expression patterns of mab21l1 and mab21l2, and conserved promoter regulation without sequence conservation. *Developmental Dynamics* **240**, 745-754
178. Yamada, R., Mizutani-Koseki, Y., Hasegawa, T., Osumi, N., Koseki, H., and Takahashi, N. (2003) Cell-autonomous involvement of Mab21l1 is essential for lens placode development. *Development* **130**, 1759-1770
179. Yamada, R., Mizutani-Koseki, Y., Koseki, H., and Takahashi, N. (2004) Requirement for Mab21l2 during development of murine retina and ventral body wall. *Developmental biology* **274**, 295-307
180. Kennedy, B. N., Stearns, G. W., Smyth, V. A., Ramamurthy, V., van Eeden, F., Ankoudinova, I., Raible, D., Hurley, J. B., and Brockerhoff, S. E. (2004) Zebrafish rx3 and mab21l2 are required during eye morphogenesis. *Developmental biology* **270**, 336-349
181. Li, S., Yin, M., Liu, S., Chen, Y., Yin, Y., Liu, T., and Zhou, J. (2010) Expression of ventral diencephalon-enriched genes in zebrafish. *Developmental Dynamics* **239**, 3368-3379
182. Wong, R. L. Y., and Chow, K. L. (2002) Depletion of Mab21l1 and Mab21l2 messages in mouse embryo arrests axial turning, and impairs notochord and neural tube differentiation. *Teratology* **65**, 70-77
183. Saito, Y., Kojima, T., and Takahashi, N. (2012) Mab21l2 is essential for embryonic heart and liver development. *PloS one* **7**, e32991
184. Wolf, L. V., Yang, Y., Wang, J., Xie, Q., Braunger, B., Tamm, E. R., Zavadil, J., and Cvekl, A. (2009) Identification of pax6-dependent gene regulatory networks in the mouse lens. *PLoS One* **4**, e4159
185. Kawahara, A., Chien, C.-B., and Dawid, I. B. (2002) The homeobox gene mbx is involved in eye and tectum development. *Developmental biology* **248**, 107-117
186. Kim, Y. J., Kim, B. G., Lee, S. J., Lee, H. K., Lee, S. H., Ryoo, H. M., and Cho, J. Y. (2007) The suppressive effect of myeloid elf-1-like factor (MEF) in osteogenic differentiation. *Journal of cellular physiology* **211**, 253-260
187. Kim, B. G., Park, Y. J., Libermann, T. A., and Cho, J. Y. (2011) PTH regulates myleoid ELF-1-like factor (MEF)-induced MAB-21-like-1 (MAB21L1) expression through the JNK1 pathway. *Journal of cellular biochemistry* **112**, 2051-2061
188. Baldessari, D., Badaloni, A., Longhi, R., Zappavigna, V., and Consalez, G. G. (2004) MAB21L2, a vertebrate member of the Male-abnormal 21 family, modulates BMP signaling and interacts with SMAD1. *BMC cell biology* **5**, 48
189. Sridharan, J., Haremaki, T., Jin, Y., Teegala, S., and Weinstein, D. C. (2012) Xmab21l3 mediates dorsoventral patterning in Xenopus laevis. *Mechanisms of development* **129**, 136-146
190. Takahashi, C., Kusakabe, M., Suzuki, T., Miyatake, K., and Nishida, E. (2015) mab21-l3 regulates cell fate specification of multiciliate cells and ionocytes. *Nature communications* **6**
191. Rainger, J., Pehlivan, D., Johansson, S., Bengani, H., Sanchez-Pulido, L., Williamson, K. A., Ture, M., Barker, H., Rosendahl, K., and Spranger, J. (2014) Monoallelic and biallelic

mutations in MAB21L2 cause a spectrum of major eye malformations. *The American Journal of Human Genetics* **94**, 915-923

192. Hartmann, R., Justesen, J., Sarkar, S. N., Sen, G. C., and Yee, V. C. (2003) Crystal structure of the 2′-specific and double-stranded RNA-activated interferon-induced antiviral protein 2′-5′-oligoadenylate synthetase. *Molecular cell* **12**, 1173-1185
193. Horn, D., Prescott, T., Houge, G., Brække, K., Rosendahl, K., Nishimura, G., FitzPatrick, D. R., and Spranger, J. (2015) A Novel Oculo-Skeletal syndrome with intellectual disability caused by a particular MAB21L2 mutation. *European journal of medical genetics* **58**, 387-391
194. Deml, B., Kariminejad, A., Borujerdi, R. H., Muheisen, S., Reis, L. M., and Semina, E. V. (2015) Mutations in MAB21L2 result in ocular coloboma, microcornea and cataracts. *PLoS Genet* **11**, e1005002
195. Lu, C., Xu, H., Ranjith-Kumar, C., Brooks, M. T., Hou, T. Y., Hu, F., Herr, A. B., Strong, R. K., Kao, C. C., and Li, P. (2010) The structural basis of 5′ triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. *Structure* **18**, 1032-1043
196. Cui, S., Eisenächer, K., Kirchhofer, A., Brzózka, K., Lammens, A., Lammens, K., Fujita, T., Conzelmann, K.-K., Krug, A., and Hopfner, K.-P. (2008) The C-terminal regulatory domain is the RNA 5′-triphosphate sensor of RIG-I. *Molecular cell* **29**, 169-179
197. Lu, C., Ranjith-Kumar, C., Hao, L., Kao, C. C., and Li, P. (2011) Crystal structure of RIG-I C-terminal domain bound to blunt-ended double-strand RNA without 5′ triphosphate. *Nucleic acids research* **39**, 1565-1575
198. Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., and Tuschl, T. (2010) Structural and functional insights into 5 [prime]-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nature structural & molecular biology* **17**, 781-787
199. Sun, Z., Ren, H., Liu, Y., Teeling, J. L., and Gu, J. (2011) Phosphorylation of RIG-I by casein kinase II inhibits its antiviral response. *Journal of virology* **85**, 1036-1047
200. Dürr, H., Körner, C., Müller, M., Hickmann, V., and Hopfner, K.-P. (2005) X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363-373
201. Zheng, J., Yong, H. Y., Panutdaporn, N., Liu, C., Tang, K., and Luo, D. (2015) High-resolution HDX-MS reveals distinct mechanisms of RNA recognition and activation by RIG-I and MDA5. *Nucleic acids research* **43**, 1216-1230
202. Wu, B., Peisley, A., Richards, C., Yao, H., Zeng, X., Lin, C., Chu, F., Walz, T., and Hur, S. (2013) Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. *Cell* **152**, 276-289
203. Ahmad, S., and Hur, S. (2015) Helicases in Antiviral Immunity: Dual Properties as Sensors and Effectors. *Trends in biochemical sciences* **40**, 576-585
204. Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23-50
205. Abdel-Monem, M., Lauppe, H.-F., Kartenbeck, J., Dürwald, H., and Hoffmann-Berling, H. (1977) Enzymatic unwinding of DNA: III. Mode of action of escherichia coli DNA unwinding enzyme. *Journal of molecular biology* **110**, 667-685
206. Fairman-Williams, M. E., Guenther, U.-P., and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Current opinion in structural biology* **20**, 313-324
207. Myong, S., Cui, S., Cornish, P. V., Kirchhofer, A., Gack, M. U., Jung, J. U., Hopfner, K.-P., and Ha, T. (2009) Cytosolic Viral Sensor RIG-I Is a 5'-Triphosphate–Dependent Translocase on Double-Stranded RNA. *Science* **323**, 1070-1074

208. Takahasi, K., Yoneyama, M., Nishihori, T., Hirai, R., Kumeta, H., Narita, R., Gale, M., Inagaki, F., and Fujita, T. (2008) Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Molecular cell* **29**, 428-440
209. Binder, M., Eberle, F., Seitz, S., Mücke, N., Hüber, C. M., Kiani, N., Kaderali, L., Lohmann, V., Dalpke, A., and Bartenschlager, R. (2011) Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid-inducible gene-I (RIG-I). *Journal of Biological Chemistry* **286**, 27278-27287
210. Peisley, A., Wu, B., Yao, H., Walz, T., and Hur, S. (2013) RIG-I forms signaling-competent filaments in an ATP-dependent, ubiquitin-independent manner. *Molecular cell* **51**, 573-583
211. Patel, J. R., Jain, A., Chou, Y. y., Baum, A., Ha, T., and García-Sastre, A. (2013) ATPase-driven oligomerization of RIG-I on RNA allows optimal activation of type-I interferon. *EMBO reports* **14**, 780-787
212. Peisley, A., Lin, C., Wu, B., Orme-Johnson, M., Liu, M., Walz, T., and Hur, S. (2011) Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition. *Proceedings of the National Academy of Sciences* **108**, 21010-21015
213. Funabiki, M., Kato, H., Miyachi, Y., Toki, H., Motegi, H., Inoue, M., Minowa, O., Yoshida, A., Deguchi, K., and Sato, H. (2014) Autoimmune disorders associated with gain of function of the intracellular sensor MDA5. *Immunity* **40**, 199-212
214. Jang, M.-A., Kim, E. K., Nguyen, N. T., Kim, W.-J., Yoo, J.-Y., Lee, J., Jeong, Y.-M., Kim, C.-H., Kim, O.-H., and Sohn, S. (2015) Mutations in DDX58, which encodes RIG-I, cause atypical Singleton-Merten syndrome. *The American Journal of Human Genetics* **96**, 266-274
215. Rice, G. I., del Toro Duany, Y., Jenkinson, E. M., Forte, G. M., Anderson, B. H., Ariaudo, G., Bader-Meunier, B., Baildam, E. M., Battini, R., and Beresford, M. W. (2014) Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nature genetics* **46**, 503-509
216. Rutsch, F., MacDougall, M., Lu, C., Buers, I., Mamaeva, O., Nitschke, Y., Rice, G. I., Erlandsen, H., Kehl, H. G., and Thiele, H. (2015) A specific IFIH1 gain-of-function mutation causes Singleton-Merten syndrome. *The American Journal of Human Genetics* **96**, 275-282
217. Louber, J., Brunel, J., Uchikawa, E., Cusack, S., and Gerlier, D. (2015) Kinetic discrimination of self/non-self RNA by the ATPase activity of RIG-I and MDA5. *BMC biology* **13**, 54
218. Lässig, C., Matheisl, S., Sparrer, K. M., de Oliveira Mann, C. C., Moldt, M., Patel, J. R., Goldeck, M., Hartmann, G., García-Sastre, A., and Hornung, V. (2016) ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *eLife* **4**, e10859
219. Rawling, D. C., Fitzgerald, M. E., and Pyle, A. M. (2015) Establishing the role of ATP for the function of the RIG-I innate immune sensor. *Elife* **4**, e09391
220. Anger, A. M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., and Beckmann, R. (2013) Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80-85
221. Herzner, A.-M., Hagmann, C. A., Goldeck, M., Wolter, S., Kübler, K., Wittmann, S., Gramberg, T., Andreeva, L., Hopfner, K.-P., and Mertens, C. (2015) Sequence-specific activation of the DNA sensor cGAS by Y-form DNA structures as found in primary HIV-1 cDNA. *Nature immunology* **16**, 1025-1033
222. Yoh, S. M., Schneider, M., Seifried, J., Soonthornvacharin, S., Akleh, R. E., Olivieri, K. C., De Jesus, P. D., Ruan, C., de Castro, E., and Ruiz, P. A. (2015) PQBP1 is a proximal sensor of the cGAS-dependent innate response to HIV-1. *Cell* **161**, 1293-1305

223. Gao, D., Li, T., Li, X.-D., Chen, X., Li, Q.-Z., Wight-Carter, M., and Chen, Z. J. (2015) Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases. *Proceedings of the National Academy of Sciences* **112**, E5699-E5705
224. Torralba, S., Sojat, J., and Hartmann, R. (2008) 2'-5'oligoadenylate synthetase shares active site architecture with the archaeal CCA-adding enzyme. *Cellular and molecular life sciences* **65**, 2613-2620
225. Balbo, P. B., and Bohm, A. (2007) Mechanism of poly (A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. *Structure* **15**, 1117-1131
226. Wu, X., Wu, F.-H., Wang, X., Wang, L., Siedow, J. N., Zhang, W., and Pei, Z.-M. (2014) Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic acids research* **42**, 8243-8257
227. Xiong, Y., and Steitz, T. A. (2006) A story with a good ending: tRNA 3′-end maturation by CCA-adding enzymes. *Current opinion in structural biology* **16**, 12-17
228. Davies, B. W., Bogard, R. W., Young, T. S., and Mekalanos, J. J. (2012) Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. *Cell* **149**, 358-370
229. Kranzusch, P. J., Lee, A. S., Wilson, S. C., Solovykh, M. S., Vance, R. E., Berger, J. M., and Doudna, J. A. (2014) Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* **158**, 1011-1021
230. Kato, K., Ishii, R., Hirano, S., Ishitani, R., and Nureki, O. (2015) Structural Basis for the Catalytic Mechanism of DncV, Bacterial Homolog of Cyclic GMP-AMP Synthase. *Structure* **23**, 843-850
231. Ming, Z., Wang, W., Xie, Y., Ding, P., Chen, Y., Jin, D., Sun, Y., Xia, B., Yan, L., and Lou, Z. (2014) Crystal structure of the novel di-nucleotide cyclase from Vibrio cholerae (DncV) responsible for synthesizing a hybrid cyclic GMP-AMP. *Cell research* **24**, 1270
232. Zhu, D., Wang, L., Shang, G., Liu, X., Zhu, J., Lu, D., Wang, L., Kan, B., Zhang, J.-r., and Xiang, Y. (2014) Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. *Molecular cell* **55**, 931-937
233. Kranzusch, P. J., Wilson, S. C., Lee, A. S., Berger, J. M., Doudna, J. A., and Vance, R. E. (2015) Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2′, 3′ cGAMP Signaling. *Molecular cell* **59**, 891-903
234. Gao, J., Tao, J., Liang, W., Zhao, M., Du, X., Cui, S., Duan, H., Kan, B., Su, X., and Jiang, Z. (2015) Identification and characterization of phosphodiesterases that specifically degrade 3′ 3′-cyclic GMP-AMP. *Cell research* **25**, 539-550
235. Zhao, G., Shi, L., Qiu, D., Hu, H., and Kao, P. N. (2005) NF45/ILF2 tissue expression, promoter analysis, and interleukin-2 transactivating function. *Experimental cell research* **305**, 312-323
236. Wolkowicz, U. M., and Cook, A. G. (2012) NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic acids research* **40**, 9356-9368
237. Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., Hatano, E., Fukushima, A., Taniguchi, T., and Agata, Y. (2009) The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Molecular and cellular biology* **29**, 3754-3769
238. Skalicky, S. E., White, A. J., Grigg, J. R., Martin, F., Smith, J., Jones, M., Donaldson, C., Smith, J. E., Flaherty, M., and Jamieson, R. V. (2013) Microphthalmia, anophthalmia, and coloboma and associated ocular and systemic features: understanding the spectrum. *JAMA ophthalmology* **131**, 1517-1524

REFERENCES

239. Fantes, J., Ragge, N. K., Lynch, S.-A., McGill, N. I., Collin, J. R. O., Howard-Peebles, P. N., Hayward, C., Vivian, A. J., Williamson, K., and van Heyningen, V. (2003) Mutations in SOX2 cause anophthalmia. *Nature genetics* **33**, 462-463
240. Ragge, N. K., Brown, A. G., Poloschek, C. M., Lorenz, B., Henderson, R. A., Clarke, M. P., Russell-Eggitt, I., Fielder, A., Gerrelli, D., and Martinez-Barbera, J. P. (2005) Heterozygous mutations of OTX2 cause severe ocular malformations. *The American Journal of Human Genetics* **76**, 1008-1022
241. Glaser, T., Jepeal, L., Edwards, J. G., Young, S. R., Favor, J., and Maas, R. L. (1994) PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nature genetics* **7**, 463-471
242. Cibois, M., Scerbo, P., Thomé, V., Pasini, A., and Kodjabachian, L. (2014) Induction and differentiation of the Xenopus ciliated embryonic epidermis. *Xenopus Development*, 112-129
243. Dubaissi, E., Rousseau, K., Lea, R., Soto, X., Nardeosingh, S., Schweickert, A., Amaya, E., Thornton, D. J., and Papalopulu, N. (2014) A secretory cell type develops alongside multiciliated cells, ionocytes and goblet cells, and provides a protective, anti-infective function in the frog embryonic mucociliary epidermis. *Development* **141**, 1514-1525
244. Karp, P. H., Moninger, T. O., Pary Weber, S., Nesselhauf, T. S., Launspach, J. L., Zabner, J., and Welsh, M. J. (2002) An in vitro model of differentiated human airway epithelia. *Epithelial cell culture protocols*, 115-137
245. Cibois, M., Luxardi, G., Chevalier, B., Thomé, V., Mercey, O., Zaragosi, L.-E., Barbry, P., Pasini, A., Marcet, B., and Kodjabachian, L. (2015) BMP signalling controls the construction of vertebrate mucociliary epithelia. *Development* **142**, 2352-2363
246. Deblandre, G. A., Wettstein, D. A., Koyano-Nakagawa, N., and Kintner, C. (1999) A two-step mechanism generates the spacing pattern of the ciliated cells in the skin of Xenopus embryos. *Development* **126**, 4715-4728
247. Stubbs, J. L., Davidson, L., Keller, R., and Kintner, C. (2006) Radial intercalation of ciliated cells during Xenopus skin development. *Development* **133**, 2507-2515
248. He, L., He, X., Lim, L. P., De Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., and Ridzon, D. (2007) A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130-1134
249. Song, R., Walentek, P., Sponer, N., Klimke, A., Lee, J. S., Dixon, G., Harland, R., Wan, Y., Lishko, P., and Lize, M. (2014) miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110. *Nature* **510**, 115-120
250. Marcet, B., Chevalier, B., Luxardi, G., Coraux, C., Zaragosi, L.-E., Cibois, M., Robbe-Sermesant, K., Jolly, T., Cardinaud, B., and Moreilhon, C. (2011) Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nature cell biology* **13**, 693-699
251. Lizé, M., Herr, C., Klimke, A., Bals, R., and Dobbelstein, M. (2010) MicroRNA-449a levels increase by several orders of magnitude during mucociliary differentiation of airway epithelia. *Cell Cycle* **9**, 4579-4583
252. Lizé, M., Klimke, A., and Dobbelstein, M. (2011) MicroRNA-449 in cell fate determination. *Cell cycle* **10**, 2874-2882
253. Park, T. J., Haigo, S. L., and Wallingford, J. B. (2006) Ciliogenesis defects in embryos lacking inturned or fuzzy function are associated with failure of planar cell polarity and Hedgehog signaling. *Nature genetics* **38**, 303-311
254. Wallingford, J. B. (2006) Planar cell polarity, ciliogenesis and neural tube defects. *Human molecular genetics* **15**, R227-R234

255. Zohn, I. E., Anderson, K. V., and Niswander, L. (2005) Using genomewide mutagenesis screens to identify the genes required for neural tube closure in the mouse. *Birth Defects Research Part A: Clinical and Molecular Teratology* **73**, 583-590
256. Davis, B. N., Hilyard, A. C., Nguyen, P. H., Lagna, G., and Hata, A. (2010) Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. *Molecular cell* **39**, 373-384
257. Saj, A., and Lai, E. C. (2011) Control of microRNA biogenesis and transcription by cell signaling pathways. *Current opinion in genetics & development* **21**, 504-510
258. Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., Li, J. B., Seeburg, P. H., and Walkley, C. R. (2015) RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115-1120
259. Rawling, D. C., Kohlway, A. S., Luo, D., Ding, S. C., and Pyle, A. M. (2014) The RIG-I ATPase core has evolved a functional requirement for allosteric stabilization by the Pincer domain. *Nucleic acids research*, gku817
260. Kohlway, A., Luo, D., Rawling, D. C., Ding, S. C., and Pyle, A. M. (2013) Defining the functional determinants for RNA surveillance by RIG-I. *EMBO reports* **14**, 772-779
261. Ramanathan, A., Devarkar, S. C., Jiang, F., Miller, M. T., Khan, A. G., Marcotrigiano, J., and Patel, S. S. (2015) The autoinhibitory CARD2-Hel2i Interface of RIG-I governs RNA selection. *Nucleic acids research*, gkv1299
262. Wu, B., and Hur, S. (2015) How RIG-I like receptors activate MAVS. *Current opinion in virology* **12**, 91-98
263. Peisley, A., Wu, B., Xu, H., Chen, Z. J., and Hur, S. (2014) Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I. *Nature* **509**, 110-114
264. Bruns, A. M., Pollpeter, D., Hadizadeh, N., Myong, S., Marko, J. F., and Horvath, C. M. (2013) ATP hydrolysis enhances RNA recognition and antiviral signal transduction by the innate immune sensor, laboratory of genetics and physiology 2 (LGP2). *Journal of Biological Chemistry* **288**, 938-946
265. Yao, H., Dittmann, M., Peisley, A., Hoffmann, H.-H., Gilmore, R. H., Schmidt, T., Schmid-Burgk, J. L., Hornung, V., Rice, C. M., and Hur, S. (2015) ATP-dependent effector-like functions of RIG-I-like receptors. *Molecular cell* **58**, 541-548
266. Weber, M., Sediri, H., Felgenhauer, U., Binzen, I., Bänfer, S., Jacob, R., Brunotte, L., García-Sastre, A., Schmid-Burgk, J. L., and Schmidt, T. (2015) Influenza virus adaptation PB2-627K modulates nucleocapsid inhibition by the pathogen sensor RIG-I. *Cell host & microbe* **17**, 309-319
267. Sato, S., Li, K., Kameyama, T., Hayashi, T., Ishida, Y., Murakami, S., Watanabe, T., Iijima, S., Sakurai, Y., and Watashi, K. (2015) The RNA sensor RIG-I dually functions as an innate sensor and direct antiviral factor for hepatitis B virus. *Immunity* **42**, 123-132
268. Zhang, H.-X., Liu, Z.-X., Sun, Y.-P., Zhu, J., Lu, S.-Y., Liu, X.-S., Huang, Q.-H., Xie, Y.-Y., Zhu, H.-B., and Dang, S.-Y. (2013) Rig-I regulates NF-κB activity through binding to Nf-κb1 3′-UTR mRNA. *Proceedings of the National Academy of Sciences* **110**, 6459-6464
269. Wilson, R. C., and Doudna, J. A. (2013) Molecular mechanisms of RNA interference. *Annual review of biophysics* **42**, 217-239
270. Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A. A., Stern-Ginossar, N., Yakhini, Z., Weinberger, A., and Segal, E. (2016) Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, aad4939
271. Kossinova, O., Malygin, A., Krol, A., and Karpova, G. (2014) The SBP2 protein central to selenoprotein synthesis contacts the human ribosome at expansion segment 7L of the 28S rRNA. *RNA* **20**, 1046-1056

272. Ablasser, A., Hertrich, C., Waßermann, R., and Hornung, V. (2013) Nucleic acid driven sterile inflammation. *Clinical Immunology* **147**, 207-215
273. Kato, H., and Fujita, T. (2015) RIG-I-like receptors and autoimmune diseases. *Current opinion in immunology* **37**, 40-45
274. Crow, Y. J., Hayward, B. E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D. N., van Bokhoven, H., Brunner, H. G., and Hamel, B. C. (2006) Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nature genetics* **38**, 917-920
275. Crow, Y. J., Leitch, A., Hayward, B. E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., and Babul-Hirji, R. (2006) Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nature genetics* **38**, 910-916
276. Rice, G. I., Bond, J., Asipu, A., Brunette, R. L., Manfield, I. W., Carr, I. M., Fuller, J. C., Jackson, R. M., Lamb, T., and Briggs, T. A. (2009) Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. *Nature genetics* **41**, 829-832
277. Rice, G. I., Kasher, P. R., Forte, G. M., Mannion, N. M., Greenwood, S. M., Szynkiewicz, M., Dickerson, J. E., Bhaskar, S. S., Zampini, M., and Briggs, T. A. (2012) Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nature genetics* **44**, 1243-1248
278. Oda, H., Nakagawa, K., Abe, J., Awaya, T., Funabiki, M., Hijikata, A., Nishikomori, R., Funatsuka, M., Ohshima, Y., and Sugawara, Y. (2014) Aicardi-Goutieres syndrome is caused by IFIH1 mutations. *The American Journal of Human Genetics* **95**, 121-125
279. An, J., Woodward, J. J., Sasaki, T., Minie, M., and Elkon, K. B. (2015) Cutting Edge: Antimalarial Drugs Inhibit IFN-β Production through Blockade of Cyclic GMP-AMP Synthase–DNA Interaction. *The Journal of Immunology* **194**, 4089-4093
280. Ablasser, A., Hemmerling, I., Schmid-Burgk, J. L., Behrendt, R., Roers, A., and Hornung, V. (2014) TREX1 deficiency triggers cell-autonomous immunity in a cGAS-dependent manner. *The Journal of Immunology* **192**, 5993-5997
281. Sjöstrand, M., Brauner, S., Kvarnström, M., Wahren-Herlenius, M., and Espinosa, A. (2012) TRIM genes are part of the interferon signature observed in patients with primary Sjögren's syndrome. *Annals of the Rheumatic Diseases* **71**, A81-A81
282. Liu, X., Jiao, Y., Wen, X., Wang, L., Ma, C., Gao, X., Chen, Z.-J., and Zhao, Y. (2011) Possible association of VISA gene polymorphisms with susceptibility to systemic lupus erythematosus in Chinese population. *Molecular biology reports* **38**, 4583-4588
283. Li, L., Yin, Q., Kuss, P., Maliga, Z., Millán, J. L., Wu, H., and Mitchison, T. J. (2014) Hydrolysis of 2′ 3′-cGAMP by ENPP1 and design of nonhydrolyzable analogs. *Nature chemical biology* **10**, 1043-1048
284. Kumar, H., Koyama, S., Ishii, K. J., Kawai, T., and Akira, S. (2008) Cutting edge: cooperation of IPS-1-and TRIF-dependent pathways in poly IC-enhanced antibody production and cytotoxic T cell responses. *The Journal of Immunology* **180**, 683-687
285. Karaolis, D. K., Means, T. K., Yang, D., Takahashi, M., Yoshimura, T., Muraille, E., Philpott, D., Schroeder, J. T., Hyodo, M., and Hayakawa, Y. (2007) Bacterial c-di-GMP is an immunostimulatory molecule. *The Journal of Immunology* **178**, 2171-2181
286. McWhirter, S. M., Barbalat, R., Monroe, K. M., Fontana, M. F., Hyodo, M., Joncker, N. T., Ishii, K. J., Akira, S., Colonna, M., and Chen, Z. J. (2009) A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. *The Journal of experimental medicine* **206**, 1899-1911

287. Škrnjug, I., Rueckert, C., Libanova, R., Lienenklaus, S., Weiss, S., and Guzmán, C. A. (2014) The mucosal adjuvant cyclic di-AMP exerts immune stimulatory effects on dendritic cells and macrophages. *PloS one* **9**, e95728

REFERENCES

Acknowledgements

Foremost I would like to thank my supervisor Prof. Dr. Karl-Peter Hopfner for the opportunity to work in his research group in the field of structural biology. I am thankful for his support, advice and especially for the freedom to develop my own ideas, interest and skills during my PhD.

Special thanks go to Dr. Gregor Witte for his support, contributions to the project and for introducing me into crystallography, along with its patience during the endless hours at the synchrotron measuring MAB21 protein crystals. You made synchrotron trips enjoyable! Alongside Gregor, I would like to thank Dr. Katja Lammens and Dr. Dirk Kostrewa for help and for answering all my crystallography related questions.

My dear lab mates in the innate immune field thank you for the great lab atmosphere - along with your scientific and personal support.

I would like to thank the whole Hopfner lab for the inspiring scientific discussions and help in the lab.

Furthermore, I would like to thank the Coordinator Dr. Petra Runge-Wollmann and the members of the graduate program GRK1721 “Integrated Analysis of Macromolecular Complexes and Hybrid Methods in Genome Biology” for support and the inspiring discussions, workshops and retreats.

My greatest gratitude goes to my family and Philipp who have encouraged me in my ambition and are always there for me.