

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

Structural and Functional Characterization of Pattern Recognition Receptors of the Innate Immune System



Diana Angela Pippig

aus

Plauen

München, 2010

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Karl-Peter Hopfner betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfsmittel erarbeitet.

München, am 17.08.2010

.....
Diana Pippig

Dissertation eingereicht am: 17.08.2010

1. Gutachter: Prof. Dr. Karl-Peter Hopfner

2. Gutachter: Prof. Dr. Elena Conti

Mündliche Prüfung am: 19.10.2010

This thesis has been prepared from February 2007 to August 2010 in the laboratory of Professor Dr. Karl-Peter Hopfner at the Gene Center of the Ludwig-Maximilians-University of Munich (LMU).

Parts of this thesis have been published:

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 Res** 37(6): 2014-2025.

Parts of this thesis have been presented at the following international conferences:

Posters displaying the "Structural and Functional Characterization of RIG-I Like Receptors" were presented at:

RNA 2008: Thirteenth Annual Meeting of the RNA Society; Berlin, Germany, August 2008

EMBO Conference Series: Helicases and NTP-Driven Nucleic Acid Motors – Structure, Function, Mechanism and Roles in Human Disease; Les Diablerets, Switzerland, June 2009

„Dass ich erkenne, was die Welt im Innersten zusammenhält“

Goethe, Faust I

1. Introduction	1
1.1. <i>The Immune System</i>	1
1.1.1. <i>Surface Barriers as First Primitive Stage of Immune Defense</i>	1
1.1.2. <i>Innate Immunity – The Second Stage</i>	2
1.1.3. <i>Adaptive Immunity – A Third Stage in Vertebrate Immunity</i>	3
1.2. <i>Pattern Recognition Receptors of the Innate Immune System</i>	4
1.2.1. <i>Nucleic Acid Responsive PRRs</i>	6
1.2.2. <i>PRR’s Knowing Friend from Foe</i>	8
1.3. <i>RIG-I-like Receptors</i>	9
1.3.1. <i>LGP2 – The Odd Member of the RLR Family</i>	11
1.3.2. <i>RD – Regulatory or Repressor Domain?</i>	11
1.4. <i>Inflammasomes – Stress and Infection Inducible Multi Protein Platforms</i>	14
1.4.1. <i>Types of Inflammasomes</i>	14
1.4.2. <i>RLR Signaling and Inflammasomes – a Possible Intersection</i>	16
1.4.3. <i>The AIM2 Inflammasome – a Cytosolic DNA sensor</i>	16
1.4.4. <i>AIM2 and the Interferon-Inducible p200 Protein Family</i>	17
1.4.5. <i>ASC – A Versatile Adaptor in Inflammation and Innate Immunity</i>	19
1.5. <i>Objectives</i>	21
2. Material and Methods	22
2.1. <i>Materials</i>	22
2.1.1. <i>Chemicals</i>	22
2.1.2. <i>Media and Supplements</i>	22
2.1.3. <i>Bacterial Strains</i>	23
2.1.4. <i>Plasmids</i>	23
2.1.5. <i>Cloning and Mutagenesis Primer</i>	24
2.1.6. <i>RNA and DNA Oligonucleotides</i>	25
2.2. <i>Methods</i>	27
2.2.1. <i>Molecular Biological Methods</i>	27
2.2.1.1. <i>Molecular Cloning</i>	27
2.2.1.2. <i>Site Directed Mutagenesis</i>	28
2.2.1.3. <i>Transformation</i>	28
2.2.1.4. <i>Plasmid Preparation</i>	28
2.2.1.5. <i>Bacmid Preparation</i>	28
2.2.2. <i>Protein Biochemical Methods</i>	29
2.2.2.1. <i>Protein Expression in Insect Cells</i>	29
2.2.2.2. <i>Protein Expression in E.coli</i>	29

Table of Contents

2.2.2.3. Protein Purification	30
2.2.2.3.1. Glutathione-S-Transferase Affinity Chromatography	30
2.2.2.3.2. Nickel Affinity Chromatography	32
2.2.2.3.3. Heparin Affinity Chromatography	32
2.2.2.3.4. Dialysis.....	32
2.2.2.3.5. An- and Cation Exchange Chromatography.....	32
2.2.2.3.6. Size Exclusion Chromatography (SEC or Gelfiltration).....	33
2.2.3. Crystallographic Methods	33
2.2.3.1. Crystallization.....	33
2.2.3.2. Crystallographic Data Collection of LGP2 RD.....	34
2.2.3.3. Structure Determination of LGP2 RD	34
2.2.3.3.1. Theoretical Background.....	34
2.2.3.3.2. Solution of the LGP2 RD Structure.....	35
2.2.4. RNA and DNA Biochemistry.....	36
2.2.4.1. RNA Preparation	36
2.2.4.2. Ribozymes and DNazymes	37
2.2.5. Biochemical Assays.....	37
2.2.5.1. Fluorescence Anisotropy Measurements	37
2.2.5.2. Electrophoretic Mobility Shift Assays.....	38
2.2.5.3. Pulldown Assays	38
2.2.5.4. Western Blots and Immunostaining	39
2.2.6. Bioinformatic Methods	39
2.2.6.1. Sequence Alignments	39
2.2.6.2. Calculation of Protein Parameters.....	39
2.2.6.3. Structure Visualization and Analysis	39
2.2.6.4. Protein Profile Search	40
2.2.6.5. Structural Homology Modeling	40
2.2.6.6. Secondary Structure Predictions	40
2.2.7. Analytical Methods	40
2.2.7.1. Mass Spectrometry.....	40
2.2.7.2. Edman-Sequencing	40
3. LGP2 – Results	41
3.1. Full Length LGP2.....	41
3.2. The Regulatory Domain of LGP2.....	42
3.2.1. Constructs and Purification.....	42
3.2.2. Crystallization and Structure Determination of LGP2 RD	43

3.2.3. Overall Structure	44
3.2.4. Comparison of LGP2 RD to RIG-I and MDA5 RDs.....	47
3.2.5. LGP2 RD Binds to dsRNA in a 5'-Triphosphate Independent Manner.....	50
3.2.6. The dsRNA Binding Site of LGP2.....	52
3.2.6.1. Study of LGP2 RD's RNA Interaction by Fluorescence Anisotropy.....	53
3.2.6.2. Electrophoretic Mobility Shift Assays of LGP2 RD – RNA complexes	54
3.3. RD – RNA Complex Crystallization Attempts	55
3.3.1. Generation of 5'-Triphosphate RNAs for Co-crystallization with RIG-I RD.....	55
3.3.2. Co-crystallization of RIG-I RD with 5'-Triphosphate RNA	59
4. AIM2 – Results	61
4.1. Full Length Mouse AIM2.....	61
4.1.1. Identification and Purification of Degradation Products of mAIM2	61
4.1.2. Structural Model of AIM2.....	62
4.2. The AIM2 HIN Domain	64
4.3. Evaluation of mAIM2 – DNA Complex Formation by Electrophoretic Mobility Shift Assays	65
4.4. Analytical Gelfiltration of Complexes of mAIM2 and DNA Ligands Suited for Crystallization	66
4.5. Crystallization of AIM2 – dsDNA Complexes	67
4.6. AIM2 and ASC Interaction	69
5. Discussion	73
5.1. RLR Regulatory Domains Have a Common RNA Binding Site.....	73
5.1.1. RLR RNA Binding Specificities are Determined by a Variable Loop Region in the RD.....	75
5.1.2. LGP2 RD Binds to dsRNA Ends	75
5.2. LGP2 as a Regulator of RIG-I and MDA5 Signaling.....	78
5.3. Possible Mechanisms of LGP2 Interference with MDA5 and RIG-I Signaling	80
5.4. The AIM2 Inflammasome – Preliminary Functional Insights	84
5.4.1. Crystallization of AIM2 – DNA Complexes.....	86
5.5. Comparison of RLRs and AIM2.....	86
6. Summary.....	88
7. References.....	90

Abbreviations

Curriculum Vitae

Acknowledgements

1. **Introduction**

1.1. *The Immune System*

The immune system is a means of antagonizing pathogenic infection of and by an organism. It is, with varying complexity, present in all kingdoms of life. In higher developed organisms the immune system features a certain hierarchy of protection levels. Thereby the specificity of the defense reaction increases with the respective level reached by the pathogen attacking the host. Namely, these stages comprise first physical, chemical or biological surface barriers and secondly components of the so called innate immune system. Solely in jawed vertebrates a third, evolutionary younger, defense strategy has developed that is termed adaptive immunity.

Major tasks of the vertebrate immune system comprise the chemoattractive recruitment of immunologically active cells to infection foci and the identification and neutralization of pathogenic substances by such specialized leukocytes. Further, the removal of infectious agents and infected cells, as well as the establishment of a memory function towards the respective antigen stimulus have to be prompted (Alberts *et al.* 2002).

1.1.1. *Surface Barriers as First Primitive Stage of Immune Defense*

Surface barriers are entirely non-specific and can be physical, such as plant cuticles, insect exoskeletons, skin and membranes that bound cells. Other mechanical defense mechanisms comprise physical removal of pathogens by peristalsis or cilia movement as well as coughing and sneezing to expel pathogens. Moreover, tears, saliva and urine have a flushing effect and mucosal excretion facilitates pathogen trapping.

The latter mechanisms also include chemical barriers. Secretions commonly harbor anti-microbial proteins or enzymes, like defensins, lysozyme, phospholipase and proteases or exhibit a low pH that is growth inhibitory towards pathogens (Alberts *et al.* 2002).

Commensal bacteria represent a biological barrier by competing with other harmful bacteria for nutrients and space or by employing their own defense mechanisms by secreting toxins or altering the environmental pH.

1.1.2. Innate Immunity – The Second Stage

The second stage of the immune system is also inherited and exhibits very little specificity. This so called innate immunity is distinguished by an immediate and maximum effect triggered in response to pathogen invasion but does not feature any immune memory function.

The innate immune response is not directed against distinct pathogens but acts in a more generic way. It is activated by a broad range of common pathogen associated molecular patterns (PAMPs) and damage or danger associated molecular patterns (DAMPs) that arise from neighboring infected or damaged cells. The innate immune system comprises cellular as well as humoral, secreted extracellular components.

Phagocytic leucocytes and macrophages represent a cellular barrier. They can sense and engulf foreign matter or microbes into endosomes. These are then fused to lysosomes that harbor various enzymes, like lysozyme and proteases or reactive oxygen species (ROS) allowing for the disintegration of the pathogenic components. Cells of the innate immune system are also involved in prompting inflammation and antigen presentation to components of the specific adaptive immune system by major histocompatibility complexes (MHC) on their surface (Alberts *et al.* 2002).

Humoral components in innate immunity comprise the complement system and various pro-inflammatory or antiviral cytokines, such as interferons (INFs) and interleukins (ILs). Complement is a complex network of constitutively present plasma and membrane-associated serum proteins which induce an inflammatory and cytolytic reaction towards pathogens or damaged tissue when activated. It is implied in chemotactic attraction of phagocytic cells, membrane rupture of foreign or infected cells, opsonization and clearance of neutralized antigen-antibody complexes. Complement thereby bridges the innate and acquired immune system branches (Dunkelberger *et al.* 2010).

Cytokines are messenger molecules and a variety of cytokine receptors, either membrane associated or cytosolic, are known. Examples are Type I and II cytokine receptors, seven transmembrane helix/G-Protein coupled receptors and the Tumor Necrosis Factor Receptor (TNFR) family. Defects in these receptors give rise to immunodeficiency. Cytokines that are upregulated by the innate immune system act either as active inhibitor of for instance viral replication (INF) or passively by facilitating chemotaxis (chemokines). Interleukins also drive the proliferation and differentiation of T-, B- and hematopoietic cells or induce fever and the upregulation of acute phase proteins (ferritin, C reactive protein, complement factors etc.) as well as further cytokines as inflammatory response.

Apart from these small effector molecules, the main molecular players in innate immunity are germline encoded pattern recognition receptors (PRRs) that specifically target PAMPs. These can be viral or bacterial nucleic acids, cell wall components or microbial toxins that are either sensed by cell surface located receptors or after entering the cell by cytosolic PRRs. Furthermore DAMPs that arise from damaged, stressed or infected cells are recognized by receptors of the innate immune system. Upon sensing of pathogenic patterns, PRRs trigger a signal transduction cascade that leads to the production and upregulation of the previously mentioned humoral, pro-inflammatory molecules (Chaplin 2010).

1.1.3. Adaptive Immunity – A Third Stage in Vertebrate Immunity

If the first two immune system barriers are evaded by the pathogen, a third stage is activated in vertebrates. This immune response is termed adaptive or acquired because it specifically acts on certain pathogens or antigens and due to its enabling immunological memory processes. The adaptive system requires first activation by the innate immune response and therefore initially exhibits a lag time between pathogen infection and the ultimate reaction. In case of recurring presentation with an already known stimulant the effect is however faster, increasing and extremely efficient.

Key players in adaptive immunity are T- and B-cells, two classes of specialized lymphocytes. They harbor a particularly diverse repertoire of antigen-specific recognition receptors and ensure specific identification and elimination of pathogens. Furthermore, they facilitate adaptive immune measures that enable tailored immune responses and long-lived memory against reinfection (Dunkelberger *et al.* 2010).

As opposed to the germline encoded PRRs of the innate immune response, the antigen-specific receptors of the adaptive system gain their diversity through somatic rearrangement of gene building blocks to form intact T-cell receptor (TCR) and immunoglobulin (B-cell antigen receptor) genes. This mode of receptor assembly from a collection of hundreds of germline-encoded gene elements in turn allows for the formation of millions of different antigen receptors and ensures unique specificity for a vast variety of antigens (Bonilla *et al.* 2010; Chaplin 2010). While T-cell receptors sense antigens only in a processed form presented by MHCs, B-cells harbor receptors that recognize raw antigens. These receptors are cell membrane standing immunoglobulins. Such antibodies also occur as humoral, secreted components of the adaptive immune system. They mainly originate from B-cell derived plasma cells and are also directed against specific antigens. Antibody-antigen complex formation favors phagocytic activity and complement activation. The immunological memory

effect is thereby guaranteed by a fraction of the antibody producing cells that remain as persistent memory B-cells (Tangye *et al.* 2009; Chaplin 2010).

1.2. Pattern Recognition Receptors of the Innate Immune System

Defending the body against intruding pathogens is an intricate undertaking that requires interplay between the innate and adaptive immune systems and unambiguous distinction between pathogenic and intrinsic patterns. The innate immune system resembles a second line of defense against pathogen infection. It exhibits a broad specificity towards a wide range of germline-encoded pathogen-associated molecular patterns (PAMPs), like microbial RNA, DNA or cell wall components that can be encountered by the host-cell. PAMP detection is mediated by various inherited pattern recognition receptors (PRRs).

PRRs trigger intracellular signaling cascades that lead to transcriptional upregulation and hence amplified expression of inflammatory mediators to coordinate the abolition of pathogens and infected cells. Importantly, deviant activation of PRR pathways can cause immunodeficiency, septic shock, or induction of autoimmunity. Thus, tight regulation is required (Takeuchi *et al.* 2010).

Generally, activation of PRR signaling pathways triggers the nuclear translocation of various transcription factors, including NF- κ B, AP-1, IRFs, and C/EBP β . This leads to the production of pro-inflammatory, chemoattractive and anti-microbial cytokines by cooperative upregulation of the transcription of their target genes. Activation of some PRRs results in their processing of precursor pro-interleukins to mature active forms.

Most PRRs are themselves IFN-inducible, allowing for an extremely robust innate immune response by positive feedback regulation. PRRs include endosomal membrane and cell surface located Toll-Like Receptors (TLRs) and C-type Lectin Receptors (CLRs) as well as intracellular Nucleotide-binding and Oligomerization Domain (NOD)-Like Receptors (NLRs). Further, Retinoid acid-Inducible Gene I (RIG-I)-Like Receptors (RLRs) and other, not yet grouped receptors like the cytosolic nucleic acid sensors AIM2 (Absent In Melanoma 2) and DAI (DNA-dependent Activator of INF Regulatory Factors) have also been described. An overview of various PRRs is given in Table 1.

Table 1 Overview of PRRs, their cellular localization, ligands and the response they trigger; adapted from (Takeuchi *et al.* 2010)

PRR	Localization	Ligand	Origin of Ligand	Response
TLR				
TLR1	Membrane	Triacyl lipoprotein	Bacteria	Cytokine production
TLR2	Membrane	Lipoprotein	Bacteria, viruses, parasites, self	Cytokine production
TLR3	Endolysosome	dsRNA	Virus	INF/Cytokine production
TLR4	Membrane	LPS	Bacteria, viruses, self	Cytokine production
TLR5	Membrane	Flagellin	Bacteria	B/T-cell differentiation
TLR6	Membrane	Diacyl lipoprotein	Bacteria, viruses	Cytokine production
TLR7 (<i>hs</i> TLR8)	Endolysosome	ssRNA	Viruses, bacteria, self	INF/Cytokine production
TLR9	Endolysosome	CpG-DNA	Viruses, bacteria, protozoa, self	INF/Cytokine production
TLR10	Endolysosome	Unknown	Unknown	Cytokine production
TLR11	Membrane	Profilin-like molecule	Protozoa	Cytokine production
RLR				
RIG-I	Cytoplasm	5'triphosphate dsRNA	RNA viruses, DNA viruses	INF/Cytokine production
MDA5	Cytoplasm	Long dsRNA	RNA viruses	INF/Cytokine production
LGP2	Cytoplasm	Unknown	RNA viruses	Regulator of RIG-I/MDA5
NLR				
NOD1	Cytoplasm	iE-DAP	Bacteria	Cytokine production
NOD2	Cytoplasm	MDP	Bacteria	Cytokine production
CLR				
Dectin-1	Membrane	β -Glucan	Fungi	Cytokine production/TLR complex inhibition
Dectin-2	Membrane	β -Glucan	Fungi	Cytokine production/TLR complex inhibition
MR	Membrane	Mannose, Glucose, N-Acetylglucosamin	Bacteria, Fungi	Cytokine production
MINCLE	Membrane	SAP130	Self, fungi	Cytokine production/TLR complex inhibition
Inflammasomes				
NLRP1	Cytoplasm	Toxins, LPS, MDP, crystals, ATP	Bacteria, endogenous DAMPs	IL-1 β /18 maturation
NLRP3	Cytoplasm	Cathepsin, ROS, ATP, crystals	Bacteria, endogenous DAMPs	IL-1 β /18 maturation
NLRC4/IPAF	Cytoplasm	Flagellin	Bacteria	IL-1 β /18 maturation
NAIP5	Cytoplasm	Flagellin	Bacteria	IL-1 β /18 maturation
AIM2	Cytoplasm	dsDNA	Bacteria, DNA viruses	IL-1 β /18 maturation
others				
DAI	Cytoplasm	dsDNA	Bacteria, DNA viruses	INF/Cytokine production
HMGB1	Cytoplasm	dsDNA	Bacteria, DNA viruses	Activation of other PRRs
RNA Pol III	Cytoplasm	dsDNA	Bacteria, DNA viruses	RLR activation by RNA transcripts

1.2.1. Nucleic Acid Responsive PRRs

Viral and pathogen derived RNA is either recognized by Toll-like receptors or by RIG-I-like Receptors or Helicases (RLR or RLH). The latter are a group of cytosolic superfamily 2 (SF2) helicases comprising RIG-I, Melanoma Differentiation Associated protein 5 (MDA5) and Laboratory of Genetics and Physiology 2 (LGP2) (Kumagai *et al.* 2010). RLRs are ubiquitously expressed and even found in cells primarily involved in adaptive immunity (Kato *et al.* 2005).

On the other hand, the presence of foreign DNA in the cytosol has been shown to be sensed by DAI (Takaoka *et al.* 2007) and indirectly by NLRP3 (NOD-Like Receptor family, Pyrin domain containing 3) (Muruve *et al.* 2008). Recently, the IFN-inducible protein AIM2 has been also implicated in pathogenic DNA sensing in the cytosol. It has been shown to form a multimeric inflammasome complex upon DNA binding and by recruiting ASC (Apoptosis-associated Speck-like protein containing a CARD; also PYCARD) and caspase-1 (Burckstummer *et al.* 2009; Fernandes-Alnemri *et al.* 2009; Hornung *et al.* 2009; Roberts *et al.* 2009; Vilaysane *et al.* 2009).

Moreover, another pathogenic DNA recognition mechanism has been revealed to link to RLR signaling. RNA Polymerase III has been shown to produce DNA derived RNA intermediates that can be sensed by RIG-I in the cytosol inducing type I interferon production (Ablasser *et al.* 2009; Chiu *et al.* 2009).

The existence of PRRs and pathways responsive to exogenous or abnormal DNA has not been known for long and it is assumed that yet more remain to be discovered.

Most of the so far described PRRs are cell-type or ligand specific. The group of High Mobility Group Box (HMGB) proteins is more versatile. Originally, they had been known to be nuclear proteins regulating chromatin structure and transcription. Only recently they have been implicated in nucleic acid delivery to PRRs for detection, by acting as more universal receptors (Yanai *et al.* 2009). A schematic overview of some of the pathways of innate immunity directed against pathogenic nucleic acids is depicted in Figure 1.

Introduction

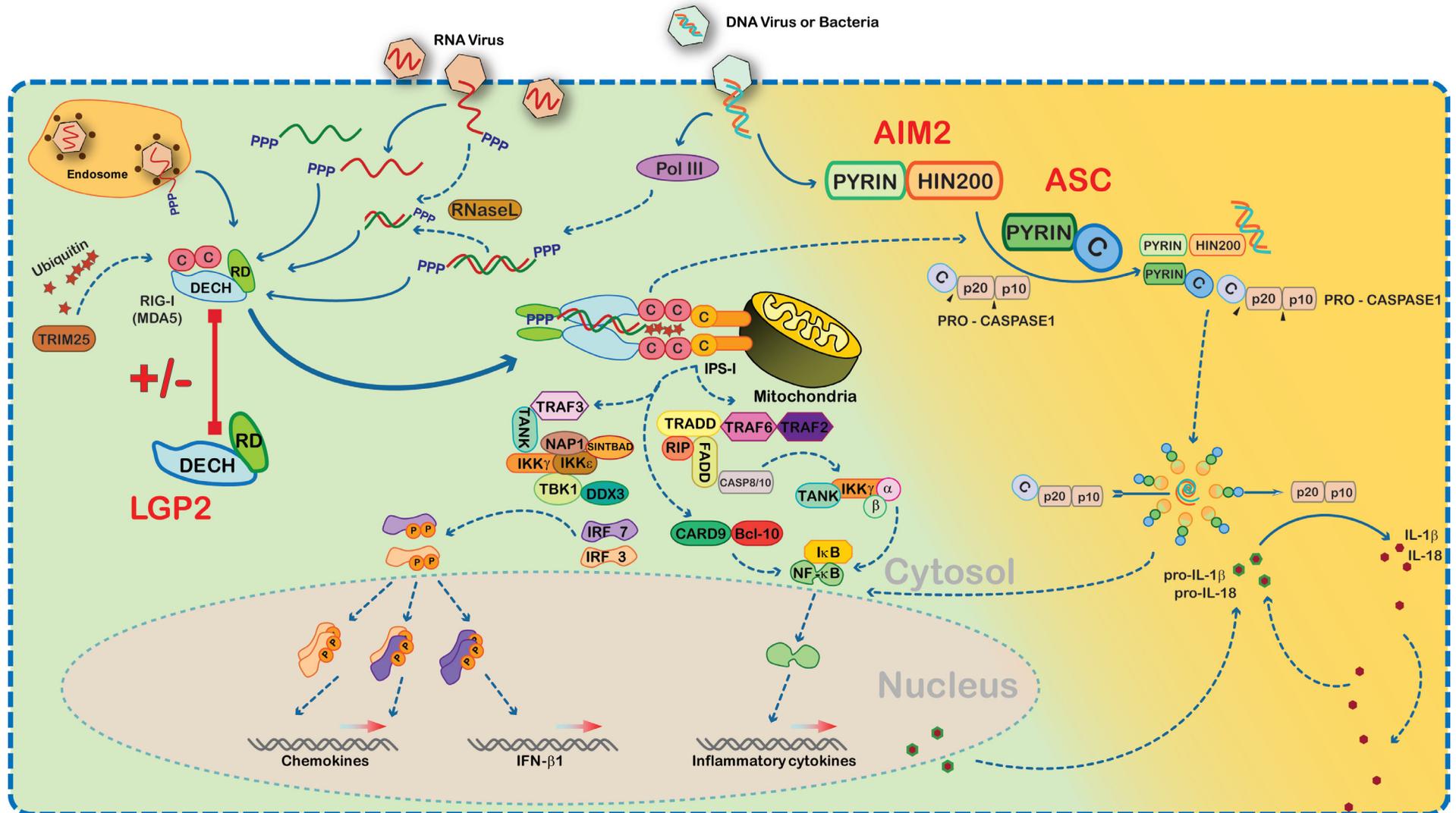


Figure 1 Schematic overview of some signaling pathways of the innate immune system directed against pathogenic nucleic acids with focus on the AIM2 inflammasome and RLR LGP2 as a regulator of RIG-I and MDA5 signaling (C = CARD).

1.2.2. PRR's Knowing Friend from Foe

PRRs stand at the beginning of a tightly regulated signaling network that ultimately triggers an antiviral or inflammatory response. Therefore discrimination between pathogenic or commensal organisms, as well as patterns occurring in the host cell itself is required. Commonly, important decisions also rely on two or more signals that are often further fine tuned by subtle thresholds for full activation of inflammatory effects and immunologically active cells.

An example of such a proofreading mechanism is that in a first instance only the expression of intracellular pro-inflammatory precursors (e.g. pro-IL-1 β and pro-IL-18) is stimulated upon PAMP recognition by PRRs. Maturation and secretion of inflammatory cytokines and hence pro-inflammatory signaling, can however be only achieved by recognition of a second “danger signal” or damage-associated molecular pattern (DAMP) in the cell. Such trigger inflammasome assembly (molecular multi protein platforms often containing NLRs or for example AIM2 and ASC as an adaptor molecule) and subsequent activation of caspase-1 that is required for the processing and release of inflammatory mediators, such as IL-1 β and IL-18. DAMPs that induce inflammasome formation can be PAMPs that are present in the cytosol, like pathogenic DNA. Further, host-cell signaling molecules that have been released by other cells suffering from stress or infection, such as reactive oxygen species (ROS) or uric acid crystals (Gallucci *et al.* 2001) act in a similar way.

Other PRRs, like RLRs, do not seem to have any “false bottom” mechanism. They can directly trigger interferon and cytokine production upon sensing pathogenic RNA in the cytosol. So the need for a different regulatory strategy arises in those pathways. This also gives rise to an important, if not the most crucial, question. How can PRRs and particularly RLRs discriminate between foreign and self patterns to, on one hand, act as extremely sensitive detectors for infection but at the same time prevent auto-immune reactions?

Generally, the mere presence of certain nucleic acid species in the cytosol is enough to trigger an immune response. Yet, it is particularly important to gain detailed understanding of what the specifically sensed patterns of each receptor are and to characterize the respective pattern receptor interaction on a molecular base (Abdul-Sater *et al.* 2009; Stutz *et al.* 2009; Latz 2010).

1.3. RIG-I-like Receptors

RLRs share a unique domain structure, consisting of a SF2 type DECH-box ATPase domain, a C-terminal regulatory domain (RD) and two N-terminal caspase activation and recruitment domains (CARDs). The latter are only found in RIG-I and MDA5, not LGP2 (Fig. 2). In addition DICER, an RNase III family member that cleaves dsRNA as well as eIF4A, that is involved in splicing, ribosome biogenesis and translation have been grouped with the other RLRs due to the high conservation of their DExD/H and HELICc (C-terminal helicase domain) motives. However, despite their common ability to bind RNA no functional relation is apparent.

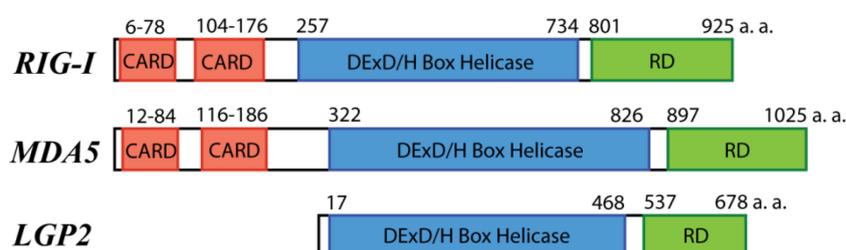


Figure 2 Domain architecture of RIG-I like Receptors.

Upon recognition of cytosolic pathogenic or pathogen-derived RNA, RIG-I and MDA5 interact with the mitochondrial membrane associated adaptor IPS-1 (Interferon- β Promoter Stimulatory protein 1; also MAVS, CARDIF and VISA) *via* a homotypic CARD domain interaction. This initiates downstream signaling and an antiviral response by interferon and cytokine production is triggered (Kawai *et al.*, 2005; Meylan *et al.*, 2005). LGP2, lacking this direct downstream interaction, is a regulator in this pathway (Yoneyama *et al.* 2005).

It has been shown, that unanchored poly-ubiquitin chains are required for CARD interaction of RIG-I and IPS-1 and therefore downstream signaling (Zeng *et al.* 2010). Upon RNA interaction RIG-I is understood to undergo a conformational change releasing the CARD domains from an auto-inhibitory state to allow for self-association and thus downstream signaling by interaction with IPS-1 (Cui *et al.* 2008). It has been speculated that the RD keeps the CARDs locked and hence RIG-I in a monomeric inactive form until it binds to specific RNA structures like 5'-triphosphates. Furthermore, LGP2's RD has been suggested to bind to RIG-I CARDs in the same manner representing a trans-inhibitory mechanism. For this reason RD has formerly also been termed Repressor Domain rather than Regulatory Domain (Saito *et al.* 2007).

Various RNA virus classes are known to be sensed by RIG-I (e.g. Rhabdo-, Paramyxo-, Orthomyxo-, Filo-, Flavi- and Reoviruses), while MDA5 so far is only understood to also sense Flavi- and Reoviruses and exclusively Picornaviruses (Kato *et al.* 2006; Loo *et al.* 2008).

The major PAMP recognized by RIG-I has been found to be 5'-triphosphate on viral RNAs. This modification arises from RNA synthesis by many viruses and is typically not found on normally capped, dephosphorylated or processed endogenous RNA molecules (Hornung *et al.* 2006). Further RIG-I stimulating species are dsRNA, 3'- and 5'-monophosphates of dsRNA, poly-U/UC rich regions in the Hepatitis C virus (HCV) genome and the synthetic dsRNA mimic polyinosine-polycytidylic acid (poly(I:C)). In addition, RNA Pol III transcribed 5'-triphosphate RNAs originating from microbial DNA entering the cell and antiviral 2'-5' oligoadenylate activated RNaseL generated small dsRNAs activate RIG-I. The latter can also arise from self-RNA in order to amplify the immune response to another stimulus (Yoneyama *et al.* 2004; Yoneyama *et al.* 2005; Malathi *et al.* 2007; Saito *et al.* 2008a; Saito *et al.* 2008b; Takahashi *et al.* 2008; Ablasser *et al.* 2009; Chiu *et al.* 2009). Ligand specificity for MDA5 is less well understood. It has been suggested that it recognizes higher order RNA structures and rather long dsRNA strands, while RIG-I senses shorter ones. An extremely potent MDA5 stimulus is poly(I:C) (Kato *et al.* 2006; Kato *et al.* 2008; Pichlmair *et al.* 2009).

LGP2 has initially been shown to bind Hepatitis C virus (HCV) RNA and Poly(I:C). The latter occurs with much higher affinity than for RIG-I and MDA5, despite the lack of a direct signaling ability of LGP2 (Yoneyama *et al.* 2005; Saito *et al.* 2007). Moreover, secondarily structured RNA and dsRNA have been introduced as possible LGP2 ligands (Rothenfusser *et al.* 2005). Still, little is known about the nature of the physiological ligand and its role in RIG-I/MDA5 regulation by LGP2.

Recently, another RIG-I signaling pathway has been discovered. 5'-triphosphate RNA exposed RIG-I, but not MDA5, has been shown to interact with the adaptor ASC to trigger caspase-1-dependent inflammasome activation and thus IL-1 β production by a mechanism independent of IPS-1 or NLRPs (NLR containing a Pyrin domain). However, this interaction could not be shown *in vitro* and other binding partners might be required (Poeck *et al.* 2010).

1.3.1. LGP2 – The Odd Member of the RLR Family

LGP2 stands out in the group of RLRs. Despite its lack of the N-terminal CARD or any other signaling domains, with the conserved helicase domain and RD it still harbors the entities that justify its being termed receptor. The high resemblance of these domains to RIG-I and MDA5 together with the missing signaling link gave early rise to speculations that LGP2 is a regulator or even inhibitor of RLR signaling (Rothenfusser *et al.* 2005; Yoneyama *et al.* 2005).

LGP2 remains relatively uncharacterized. It has been shown to interact with dsRNA and *in vivo* studies revealed a repressing effect of LGP2 on RIG-I but not MDA5. The impact of LGP2 on RIG-I signaling has thereby been assigned to a possible RNA sequestration mechanism (Rothenfusser *et al.* 2005; Komuro *et al.* 2006; Saito *et al.* 2007).

On the other hand, LGP2 has been shown to interfere with the RIG-I signaling pathway yet on another, RNA independent level. Immunoprecipitation assays revealed an interaction of LGP2 with IPS-1 in the C-terminal region, spanning residues 300-540, that include the IPS-1 mitochondrial transmembrane domain. It therefore competes with the downstream mediator kinase IKK ϵ that shares the same interaction site. Importantly, a CARD-lacking RIG-I construct that might have functioned as LGP2 mimic did not behave comparably in this study. Moreover, no binding of LGP2 to the IPS-1 CARD was observed, correlating with the finding that LGP2 and CARD-mediated RIG-I binding to IPS-1 are not exclusive (Komuro *et al.* 2006).

1.3.2. RD – Regulatory or Repressor Domain?

The way viral RNA is specifically sensed and distinguished from abundant cellular RNA is not entirely understood. It has however been shown, that RNAs harboring a 5'-triphosphate, a modification arising from unprocessed viral RNA transcripts, are capable of activating ATPase activity *in vitro* and stimulate *in vivo* signaling of RIG-I (Cui *et al.*, 2008; Hornung *et al.*, 2006). This crucial interaction, even though it is likely to be only a part of the whole RNA sensing mechanism, has been assigned to the C-terminal (RD) domain of RIG-I.

The C-terminal domains of RIG-I and LGP2 were initially referred to as Repressor Domains, after finding that RIG-I constructs lacking this domain confer constitutive signaling to the interferon- β promoter, while expression of only the C-terminal domain was inhibitory. This was also shown for LGP2 RD acting in trans to RIG-I (Saito *et al.* 2007; Vitour *et al.* 2007). Recent results suggest a more regulatory function of RD rendering it to be a

Regulatory Domain. More insight into the crucial role of the C-terminal domain in nucleic acid binding and ligand specificity has been gained. Also RD became understood to target RLRs towards pathogenic RNA in the first place (Hornung *et al.* 2006; Pichlmair *et al.* 2006; Cui *et al.* 2008).

The crystal structure of the RIG-I RD (Fig. 3 A) (Cui *et al.* 2008) is related to eukaryotic GDP/GTP exchange factors of Rab-like small GTPases, e.g. MSS4 (PDB: 1HXR, 19% identity) and to Methionine sulfoxide reductase B from the bacterium *Xanthomonas campestris* (PDB: 3HCJ, 16% identity). A functional relationship is nevertheless very unlikely, given that the sequence identity for the structural analogs is insignificant. Furthermore, the relative number of positively charged residues, a prerequisite for RNA binding, is extremely small compared to RIG-I RD. However, a prominent shared feature is a Zinc coordination site, formed by four invariant cysteine residues, which is crucial for protein integrity and hence *in vivo* signaling of RIG-I (Fig. 3 B, C).

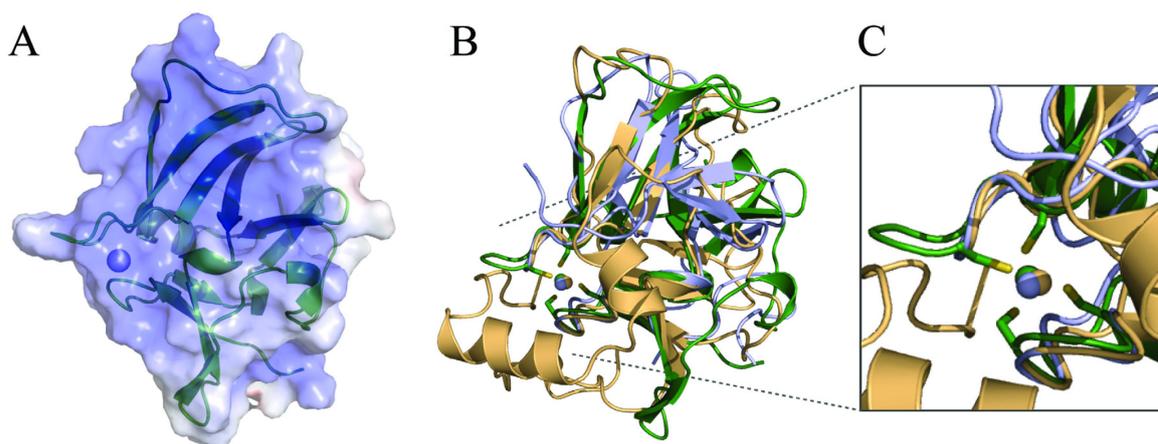


Figure 3 (A) RIG-I RD in cartoon representation with electrostatic surface charge potential (blue -8 kT to red +8 kT). (B) Superposition of RIG-I RD (green) with the two structurally related proteins nucleotide exchange factor MSS4 (1HXR, pale blue) and MsrB (3HCJ, pale orange) (Zhu *et al.* 2001; Ranaivoson *et al.* 2009). (C) Conserved Zn^{2+} coordination fold.

The RIG-I RD structure reveals an accessible, positively charged cleft that appears to be well suited to interact with the 5'-triphosphate ligand and RNA backbone phosphates (Fig. 3 A). Despite the generally high similarity of RLR RDs, several residues, shown to be crucial for 5'-triphosphate dependent binding of RNA in RIG-I RD, vary amongst the other two RLR RDs. This likely confers their prevalent selectivity against different RNA molecules (Fig. 4).

1.4. Inflammasomes – Stress and Infection Inducible Multi Protein Platforms

Inflammation is the coordinated immune response to harmful stimuli that occur due to infections or tissue damage. While it is essential for host resistance to infections, inflammation can be detrimental when produced chronically or in excess and is therefore linked to various diseases. Most notably auto-immune diseases, auto-inflammatory disorders, cancer and septic shock can result from mal-function of the inflammatory immune response. Hence, a tight regulation of inflammatory processes is indispensable (Ferrero-Miliani *et al.* 2007; Barton 2008).

In response to injurious or infectious agents caspase-1 activating cytosolic multimolecular protein complexes, termed inflammasomes, are formed (Martinon *et al.* 2002). In contrast to RLRs, inflammasomes function only in part in the transcriptional upregulation of immune response genes, but more importantly activate the cysteine protease caspase-1. The latter drives the maturation and secretion of pro-inflammatory interleukins of the IL-1 superfamily from precursors. These pro-ILs are expressed in response to other, primary PRR signaling pathways, like the RLR one. Secreted ILs are sensed by specific surface receptors on other cells.

Inflammasomes are part of the inherited immune system. They act to bridge it to the adaptive one by producing the interleukins required for stimulating B- and T-cell differentiation and antigen specific receptor production. This effect is utilized in vaccination. A commonly used adjuvant in vaccines is aluminium hydroxide, which is capable of antigen adsorption. Aluminium hydroxide activates a specific inflammasome (Hornung *et al.* 2008). The so induced IL release and therefore triggering of an adaptive immune response is specific to the introduced antigen, extremely safe and effective (Eisenbarth *et al.* 2008). Furthermore, inflammasome mediated IL release causes fever and increased acute phase protein production.

1.4.1. Types of Inflammasomes

A key player in inflammasome assembly is the adaptor protein Apoptosis-associated Speck-like protein containing a CARD (ASC). It is also termed PYCARD due to it consisting of an N-terminal Pyrin (also DAPIN: Domain in Apoptosis and INF response) and a C-terminal CARD domain. ASC is a common interaction partner in the inflammasome scaffold and usually indispensable for caspase-1 recruitment to this pro-inflammatory platform.

Despite the presence of ASC as a common adaptor, different types of inflammasomes can be distinguished (Fig. 5). A large group is made up by NOD-Like Receptor (NLR) inflammasomes. They exhibit a common domain structure usually containing a Leucine Rich Repeat (LRR), typically representing the receptor domain and a Nucleotide Binding (NBD) or NACHT (NAIP, CIITA, HET-E, TP1) domain that facilitates oligomerization upon ligand interaction. The NLR inflammasomes can be further differentiated. The NLRP (also NALP) inflammasomes additionally harbor a Pyrin domain (PYD) for ASC interaction and NLRC (also IPAF) inflammasomes lack PYD but instead contain a CARD domain for direct interaction with caspase-1. Nevertheless it has been suggested that signaling by the IPAF inflammasome is not entirely independent of ASC (Suzuki *et al.* 2007). Another NLR inflammasome is NAIP5 (also NLRB) that contains Baculoviral Inhibitor of apoptosis proteins Repeat (BIR) domain repeats instead of PYD or CARD and functions in collaboration with IPAF (Stutz *et al.* 2009; Schroder *et al.* 2010a).

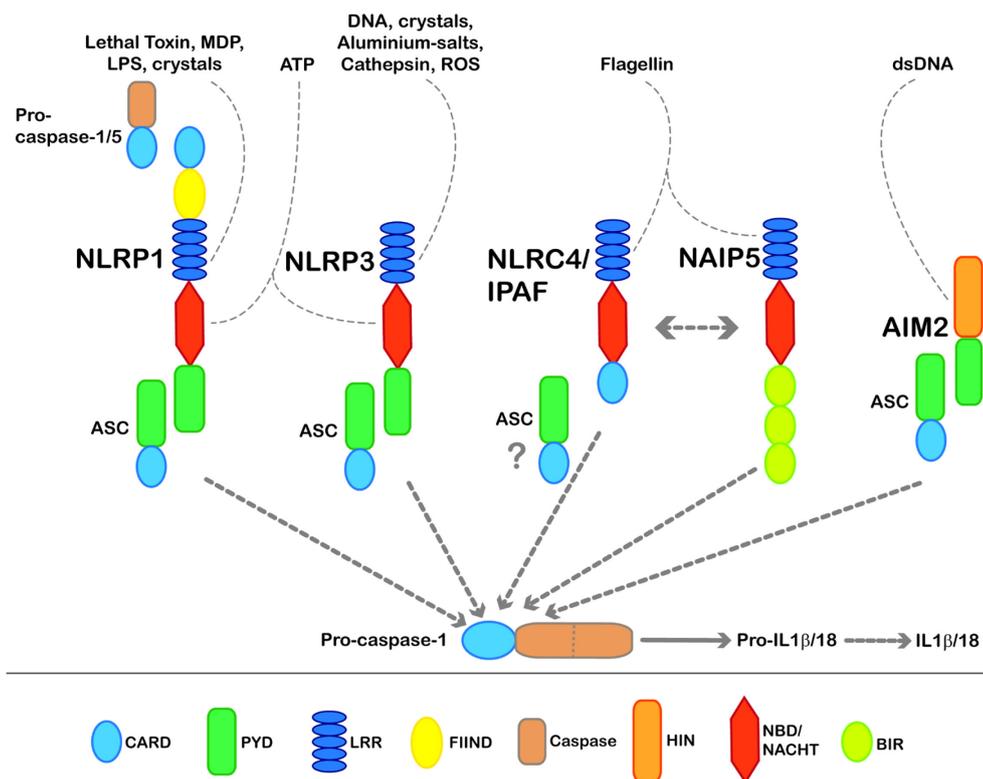


Figure 5 Overview of the assembly, domain structure and direct or indirect stimuli of different inflammasomes.

It has been shown that inflammasomes are formed by oligomeric complexes of their building blocks; however exact stoichiometries are not known for most inflammasomes. For NLRP1 penta- and heptameric assemblies have been revealed by electron microscopy (Faustin *et al.* 2007). This work suggests a donut-shaped structure for inflammasomes similar

to the structure of the human apoptosome formed by a heptameric assembly of its CARD and NOD domains (Yu *et al.* 2005).

The recently discovered AIM2 inflammasome is exceptional. It is the only so far known inflammasome that is specifically activated in direct response to cytosolic DNA. It further represents the first example of a non-NLR family member forming an inflammasome scaffold. The usual NLR motives are replaced by a C-terminal HIN-200 domain in AIM2. This part acts as a receptor of cytosolic dsDNA and is thought to confer oligomerization. AIM2 further harbors an N-terminal PYD for interaction with ASC (Burckstummer *et al.* 2009; Fernandes-Alnemri *et al.* 2009; Hornung *et al.* 2009).

In addition, the triggering of AIM2 inflammasome assembly through only one specific stimulus (dsDNA), while NLRP1 and NLRP3 are activated by various PAMPs and DAMPs, is unusual. The mechanism underlying this versatility in NLRP inflammasomes is not well understood. Evidence has arisen however, that NLRPs do not directly bind those diverse pathogenic molecules. They respond to a more unique secondary signal induced by primary PAMPs or a versatile adaptor capable of binding such (Schroder *et al.* 2010b; Tschopp *et al.* 2010).

1.4.2. RLR Signaling and Inflammasomes – a Possible Intersection

Recently RIG-I has been found to also interact with ASC upon binding to viral RNA. Therefore, it is potentially capable of forming an IL-1 β /18 processing complex comparable to the AIM2 inflammasome. Remarkably, RIG-I would thereby act in a dual role in first triggering the production of pro-inflammatory cytokines in an IPS-1 dependent manner and secondly in controlling their processing in a similar way to the inflammasome (Poeck *et al.* 2010).

1.4.3. The AIM2 Inflammasome – a Cytosolic DNA sensor

Sensors for cytoplasmic DNA have been investigated only recently. Thus, they still remain barely described and only a few have been identified. Best characterized is TLR9, which senses unmethylated CpG-rich DNA in endosomes (Chuang *et al.* 2002). Furthermore, the cytosolic PRR DAI has been shown to induce type I interferon production in response to foreign DNA (Takaoka *et al.* 2007). Also NLRP3 is implicated in capsase-1 activation specifically in response to adenoviral DNA only, likely involving another NLRP3 inflammasome activating adaptor or secondary signal (Muruve *et al.* 2008).

With AIM2, a more general sensor of cytoplasmic DNA has now been described. AIM2 initiates antiviral and inflammatory responses. It is capable of binding dsDNA with its C-terminal p200 (also HIN-200: Hematopoietic Interferon-inducible Nuclear proteins with a 200-amino-acid repeat) domain, triggering association with ASC *via* homotypic Pyrin domain interactions. Upon subsequent recruitment of pro-caspase-1 by ASC's CARD domain, complex formation is completed (Fig. 6) (Hornung *et al.* 2009).

It is also believed that this subcomplex further assembles to a large multimeric complex, the actual inflammasome. This multimerization is not entirely clear, though. There are indications that it originates from AIM2, comparable to association of NLRs *via* the NBD/NACHT domain, or ASC oligomerization. Furthermore, the assembly of the macromolecular platform could be simply mediated by clustering upon multiple binding sites on the dsDNA ligand, *via* the HIN domain of AIM2 (Fernandes-Alnemri *et al.* 2009; Hornung *et al.* 2009).

Upon inflammasome assembly pro-caspase 1 is auto-catalytically cleaved, resulting in active caspase-1 dimers. Thus processing of IL-1 β and IL-18 from precursors and their subsequent release is achieved (Burckstummer *et al.* 2009; Hornung *et al.* 2009; Vilaysane *et al.* 2009).

Furthermore, AIM2 was found to be interferon inducible and it has been shown to stimulate NF- κ B dependent reporter gene activity when overexpressed *in vivo* (Hornung *et al.* 2009).

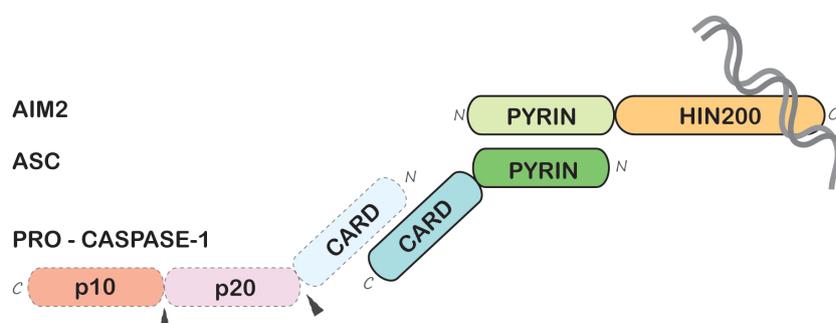


Figure 6 Schematic representations of the AIM2 inflammasome components and the way they interact upon binding of the HIN domain of AIM2 to dsDNA in the cytosol.

1.4.4. AIM2 and the Interferon-Inducible p200 Protein Family

A variety of p200-family proteins are found in human and mouse. They are encoded by IFN-inducible genes of the Ifi200 family. IFI-200 proteins were first identified as IFN-inducible nuclear proteins and implicated in cell cycle regulation and differentiation (Landolfo *et al.* 1998). This was based on their ability to interact with and modulate the

activities of multiple transcriptional factors such as pRb (Retino Blastoma protein) and p53 (Choubey *et al.* 1995; Dawson *et al.* 1996; Min *et al.* 1996; Johnstone *et al.* 2000; Ding *et al.* 2004).

In humans, four p200-family proteins have been identified so far, Interferon-Inducible protein 16 (IFI16), Myeloid Nuclear Differentiation Antigen (MNDA), IFN-inducible protein X (IFIX) and AIM2. Mice harbor some more members of this group, but AIM2 is the only real homolog between the two species (Choubey *et al.* 1995; Johnstone *et al.* 2000; Choubey *et al.* 2008; Zhang *et al.* 2009). Mouse AIM2 has been shown to function analogous to its human counterpart (Roberts *et al.* 2009).

p200 family proteins are named after at least one shared partially conserved repeat of 200 amino acid residues. This domain has been also termed HIN-200 domain (Hematopoietic Interferon-inducible Nuclear proteins with a 200-amino acid repeat), even though members have now been identified that are more ubiquitously expressed and appear cytoplasmic as well (Dawson *et al.* 1996; Roberts *et al.* 2009). HIN-200 consists of two consecutive oligonucleotide/oligosaccharide binding folds (OB-fold) that are required for DNA interaction (Albrecht *et al.* 2005). Most of the IFI-200 proteins, also AIM2, contain a Pysin domain, a common motif associated with protein-protein interactions in the regulation of apoptotic and inflammatory signaling pathways.

Furthermore, AIM2 and other p200 proteins harbor a conserved homo- or hetero-dimerization motif, “MFHATVAT”, in their HIN domains. Another well conserved putative I/LxCxE pRb binding site is found in most p200 proteins but not AIM2 (Fig. 7) (Albrecht *et al.* 2005).

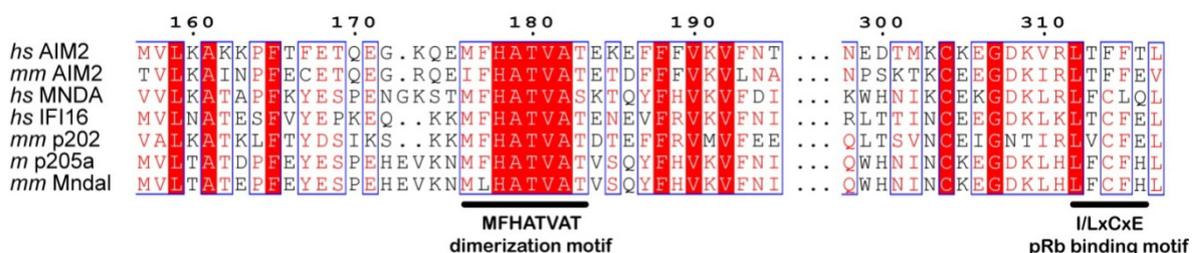


Figure 7 Multiple sequence alignment of p200 proteins of human and mouse with the homo- or hetero-dimerization motif MFHATVAT and the I/LxCxE motif, implicated in pRb binding by some group members, underlined. Identical residues are depicted in white and shaded with red, homolog residues are shown in red.

Most p200-family proteins harbor a classic nuclear localization signal (NLS) and are hence primarily detected in the nucleus. Exceptions are p202 in mouse and AIM2 in mouse and human that are preferentially cytosolic (Choubey *et al.* 2000; Ludlow *et al.* 2005).

Controversially, a nuclear localization of AIM2 has been demonstrated in one study (Cresswell *et al.* 2005). However, the NLS containing region found between the Pyrin and HIN domain in nuclear localized homologs is entirely missing in AIM2. Hence, a predominant cytosolic localization appears more reasonable. Also in a physiological context this seems more logical since AIM2 was shown to bind cytosolic dsDNA and thus inflammasome specks were also detected in the cytosol (Burckstummer *et al.* 2009; Fernandes-Alnemri *et al.* 2009; Hornung *et al.* 2009).

The HIN-200 domain of AIM2 has been shown to bind DNA, with a preference for double strands, whereas the Pyrin domain associates with the adaptor molecule ASC to activate both NF- κ B and caspase-1 (Hornung *et al.* 2009). Interestingly, mouse p202 has also been demonstrated to bind dsDNA in the cytosol, but it lacks the Pyrin domain, required for downstream interaction. Due to p202's ability to heterodimerize with AIM2 it has been suggested to be a modulator of AIM2 activity in either inflammasome formation or NF- κ B stimulation (Choubey *et al.* 2000; Roberts *et al.* 2009). For mouse p202 no human homolog has been found so far. It has however been suggested that splice variants of p200-family proteins might exist in humans that could act in a similar regulatory way (Ludlow *et al.* 2005; Lengyel *et al.* 2010).

1.4.5. ASC – A Versatile Adaptor in Inflammation and Innate Immunity

Apoptosis-associated Speck-like protein containing a CARD (ASC or PYCARD) was first identified by an antibody pulldown of insoluble components in retinoic acid exposed cells. The otherwise soluble, cytosolic 22-kDa protein exhibited intriguing behavior by forming aggregates and appeared as a speck in apoptotic cells treated with retinoic acid and other anti-tumor drugs (Masumoto *et al.* 1999). ASC harbors an N-terminal Pyrin domain (residues 1-92 for *hsASC*), a homotypic protein–protein interaction domain belonging to the six-helix bundle death domain (DD)-fold superfamily that includes DDs, death effector domains (DEDs), CARDS (Bertin *et al.* 2000; Fairbrother *et al.* 2001; Martinon *et al.* 2001; Pawlowski *et al.* 2001). Connected by a flexible linker PYD is followed by a C-terminal CARD (residues 116-195 for *hsASC*) that belongs to the same fold family (de Alba 2009). Both domains show an analogous architecture, although they have varying surface charge potentials and are in a back-to-back orientation. This acts to prevent steric interference of each domain with the binding site of the other (Fig. 8).

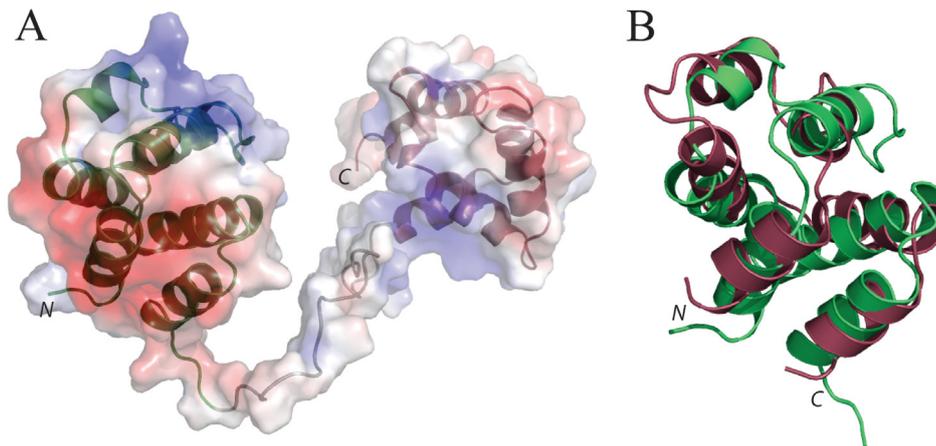


Figure 8 (A) NMR structure of full length ASC (PYCARD) with the N-terminal PYD shown in green and C-terminal CARD in purple. The electrostatic surface potential ranges from 5 kT (blue) to -5 kT (red). (B) Superposition of the six α -helix bundles of PYD and CARD of ASC.

Both domains, CARD and even more so PYD, exhibit a certain polarity, therefore accounting for self association and filament formation effects that have been reported for ASC (Masumoto *et al.* 2001; Moriya *et al.* 2005; Cheng *et al.* 2010). ASC has been shown to form dimers that subsequently oligomerize. Consistently, interaction between the PYD and CARD domains of ASC can be either hetero- or homophilic (Masumoto *et al.* 2001). Furthermore oligomerization seems pH dependent (Gattin *et al.* 2008; Cheng *et al.* 2010) and cooperative, with speck formation being an “all or none” event (Cheng *et al.* 2010).

Aside from a major part of hitherto identified inflammasomes (Fig. 5), ASC has been implicated in interaction with the pro-apoptotic protein Bax (BCL2-associated X protein) and the regulation of a p53–Bax mitochondrial pathway of apoptosis (Ohtsuka *et al.* 2004). Also binding to Pyrin and caspase-1 to form a pyroptosome has been shown (Fernandes-Alnemri *et al.* 2007; Yu *et al.* 2007). ASC was further found to interact with the viral RNA receptor RIG-I (Poeck *et al.* 2010). Interaction is generally facilitated by a PYD-PYD contact; however evidence has been provided that ASC can be also involved in IPAF inflammasome formation despite a lack of an adaptor PYD in IPAF (Geddes *et al.* 2001). The latter is likely bound *via* the CARD that is otherwise required for caspase recruitment. In addition ASC has been shown to induce adaptive immune responses independently of caspase-1 inflammasomes and to be crucial for antigen-induced T-cell priming in dendritic cells (Ippagunta *et al.* 2010).

Due to its versatility ASC is an excellent target for inflammasome regulation and modulation. Hence families of small proteins that are composed of either a CARD or a PYD only, emerged as important inflammasome regulators. These CARD-only proteins (COPs) and PYD-only proteins (POPs) function as endogenous dominant-negative proteins that

modulate activity of inflammasomes in response to pathogen infection and tissue destruction. The inhibitory effect is thereby either achieved by their binding to ASC directly or its interaction partners, in any case formation of functional inflammasomes by oligomerization of PYD and CARD containing components is disturbed (Stehlik *et al.* 2007).

1.5. Objectives

At the start of this PhD project, the field of cytosolic RLR signaling in innate immunity was only emerging. Especially little was known about the CARD-less RLR LGP2. A main goal was therefore the structural and functional dissection and characterization of this protein. This was for the purpose of gaining explanations for LGP2's regulatory behavior towards RIG-I signaling and its specific RNA interaction modes.

In this context, the general nature of RNA pattern recognition in RLRs was investigated. This was deemed to be of importance since sensitivity and specificity in ligand binding is crucial for proper RLR signaling and PRRs in general. Furthermore, the key towards understanding how RLRs discriminate between pathogenic, commensal, or intrinsic patterns lies in the molecular structure of the receptors.

During the course of this thesis, AIM2, a new DNA specific PRR had been discovered by different groups. AIM2, as a cytosolic receptor of pathogenic DNA, was considered to make for a good comparison with LGP2 and RLRs and their role as receptors of viral RNA. Furthermore AIM2 had been reported to assemble into an unusual inflammasome upon DNA interaction and little functional or structural details were known. The direct (secondary) stimuli activating other inflammasomes are unclear. Therefore, the discovered AIM2 inflammasome formation triggered by direct interaction with dsDNA can be considered as special. This makes it an ideal target for the first structural and functional analysis and *in vitro* reconstitution of a complete ligand bound inflammasomal complex. Moreover, certain intersections between the AIM2 inflammasome and RLR pathways became obvious.

For these reasons, gaining insight into the DNA induced AIM2-ASC inflammasome subcomplex assembly became another goal of this PhD project.

2. Material and Methods

2.1. Materials

2.1.1. Chemicals

All common chemicals were obtained from Roth (Karlsruhe, Germany) and Sigma (Deisenhofen, Germany), unless otherwise stated. Enzymes and nucleotides for molecular biology were supplied by MBI Fermentas (St. Leon-Rot, Germany) or New England Biolabs (Ipswich, MA, USA).

Chromatographic systems, media and columns were purchased from GE Healthcare (München, Germany). DNA oligonucleotides for cloning were ordered from Eurofins MWG (München, Germany). RNA and DNA for crystallization and assays were acquired from Biomers (Ulm, Germany) or Thermo Scientific (Ulm, Germany). Synthetic 5'PPP RNA was obtained from Eurogentec (Köln, Germany). cDNAs were received from RZPD (Heidelberg, Germany) or ImaGenes (Berlin, Germany).

Synthetic genes were obtained from Eurofins MWG (München, Germany) or Mr.Gene (Regensburg, Germany). Crystallographic tools and crystallization screens were purchased from Hampton Research (Aliso Viejo, CA, USA), Jena Bioscience (Jena, Germany), Corning (Amsterdam, Netherlands) and Qiagen (Hilden, Germany).

2.1.2. Media and Supplements

Luria Broth (LB) liquid media as well as LB Agar plates were prepared according to standard protocols (Sambrook *et al.* 1989). The media was supplemented with the respective antibiotics using stock solutions in 1:1000 dilutions (Table 2). Selenomethionine-containing protein was expressed in methionine auxotrophic *E. coli* strain B834 (Rosetta (DE3)) using LeMaster's medium supplemented with selenomethionine (Hendrickson *et al.* 1990).

Table 2 Antibiotic stock solutions

Antibiotic	Concentration (1000x)	Solvent
Ampicillin (Na-Salt)	100 mg/ml	water
Kanamycin	50 mg/ml	water
Chloramphenicol	34 mg/ml	ethanol
Tetracycline	12.5 mg/ml	ethanol
Gentamycin	10 mg/ml	water

Insect cell media powder (Express Five) was purchased from Invitrogen (Karlsruhe, Germany) and solubilized according to the manufacturer's protocol. Before use, the media was filter sterilized and supplemented with gentamycin (10 µg/ml) and glutamine (final concentration 18 mM).

2.1.3. Bacterial Strains

Table 3 Bacterial strains

<i>E.coli</i> strain	Genotype	Source
XL1 Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F'proAB lacIqZΔM15 Tn10 (TetR)]	Stratagene, Heidelberg
Rosetta (DE3)	F- <i>ompT hsdSB</i> (rB- mB-) <i>gal dcm</i> (DE3) pRARE2 (CamR)	Novagen, Madison USA
B834 / DE3	F- <i>ompT hsdSB</i> (rB- mB-) <i>gal dcm</i> <i>met</i> (DE3) pRARE2 (CamR)	Novagen, Madison USA
DH10MultiBac	not specified	Imre Berger (Berger <i>et al.</i> 2004)

2.1.4. Plasmids

Table 4 Utilized plasmids

Plasmid	Expression System	Source
pET21a(+)	<i>E. coli</i>	Novagen, Madison USA
pET28a(+)	<i>E. coli</i>	Novagen, Madison USA
pFBDM	Insect cells	Imre Berger (Berger <i>et al.</i> 2004)
pGEX6P2	<i>E. coli</i>	GE Healthcare, München, Germany
pET28M-SUMO3-GFP	<i>E. coli</i>	EMBL, Heidelberg, Germany

2.1.5. Cloning and Mutagenesis Primer

Table 5 Primer sequences used for cloning of constructs or site directed mutagenesis

primer	sequence (5' → 3')	restriction site
LGP2 537 fwd	AAAAACAT ATGGCAGCCCAGCGGAGAACCA	NdeI
LGP2 543 fwd	AAAAACAT ATGCAGCGGCAGCAGTCCCAGTG	NdeI
LGP2 671 STOP rev	AAAAAGC GGCCGCTCAGTTCTCGGCACAATG	NotI
LGP2 rev	AAAAAGC GGCCGCTCCAGGAGAGGTCCGAC	NotI
LGP2 STOP rev	AAAAAGC GGCCGCTCAGTCCAGGGAGAGGTCCGAC	NotI
LGP2 fwd	AAAAACAT ATGGAGCTTCGGTCTACCAATG	NdeI
LGP2 549 fwd	AAAAACAT ATGGTGGAGCACGTGCAGCTACTCTG	NdeI
LGP2 C615A	ATCAGCTGCAGGAACGCTGGGGAGGTCTGGGG	-
LGP2 H576Y	GGTGGAGGGCACCTACCATGTCAATGTG	-
LGP2 K634E	GCCAGTGCTCGAAGTCCGCAGCATGCTGC	-
LGP2 W604A	GTCTTCAAGGACGCGAAGCCTGGGGGTGTC	-
LGP2 K605E	CAAGGACTGGGAGCCTGGGGGTGTC	-
LGP2 P606K	GTCTTCAAGGACTGGAAGAAAGGGGTGTCATCAG	-
LGP2 L621A	GGGGAGGTCTGGGGTGCGCAGATGATCTAC	-
LGP2 K626E	GCAGATGATCTACGAGTCAGTGAAGCTGCC	-
LGP2 N583D	GTCAATGTGAACCCCGACTTCTCGAACTAC	-
RIG-I 802 fwd	AAAAACAT ATGGATAAGGAAAATAA	NdeI
RIG-I STOP rev	AAAAAGC GGCCGCTCATTTGGACATTTCTGCTG	NotI
<i>h</i> AIM2 fwd	AAAAAACAT ATGGAGAGTAAATACAAGGAGATACTCTTGC	NdeI
<i>h</i> AIM2 STOP rev	AAAAAAGC GGCCGCCTATGTTTTTTTTTTGGCCTTAATAACC	NotI
<i>h</i> AIM2 rev	AAAAAAGC GGCCGCTGTTTTTTTTTTGGCCTTAATAACC	NotI
<i>h</i> AIM2 140 fwd	AAAAAACAT ATGGCCCAGCAGGAATCTATCAGAGAA	NdeI
<i>m</i> AIM2 fwd	AAAAAACAT ATGGAGAGTGAGTACCGGGAAATG	NdeI
<i>m</i> AIM2 146 fwd	AAAAAACAT ATGGCAGAACAGGAAGCCATCAGAGA	NdeI
<i>m</i> AIM2 STOP rev	AAAAAAGC GGCCGCTCACTCCACACTTTTCATGTCAGTTTT	NotI
<i>m</i> AIM2 rev	AAAAAAGC GGCCGCCTCCACACTTTTCATGTCAGTTTT	NotI
<i>m</i> AIM2 94 fwd	AAAAAACAT ATGACCAATACAAAGAAGAG	NdeI
<i>m</i> AIM2 137 fwd	AAAAAACAT ATGGCTAAGCCTCAGAAGAAACAG	NdeI
<i>m</i> ASC 93 <i>syn</i> STOP rev	AAAAAAGC GGCCGCTTAGCCTGATTCTTCTTTGG	NotI
<i>m</i> ASC <i>syn</i> STOP rev	AAAAAAGC GGCCGCTTAGGATTGCTCCAG	NotI
<i>m</i> ASC <i>syn</i> fwd	AAAAAAG GATCCATGGGTCGTGCTCGTG	BamHI
<i>m</i> ASC <i>syn</i> fwd	AAAAAAA CCGGTATGGGTCGTGCTCGTG	AgeI
<i>m</i> ASC <i>syn</i> fwd	AAAAAACAT ATGGGTGCTGCTCGTG	NdeI
<i>h</i> ASC <i>syn</i> fwd	AAAAAAG GATCCATGGGACGTGCTCGGG	BamHI
<i>h</i> ASC <i>syn</i> fwd	AAAAAAA CCGGTATGGGACGTGCTCGGG	AgeI
<i>h</i> ASC <i>syn</i> fwd	AAAAAACAT ATGGGACGTGCTCGGG	NdeI
<i>h</i> ASC fwd	AAAAAACAT ATGGGGCGCGCGCGACGCC	NdeI
<i>h</i> ASC STOP rev	AAAAAAGC GGCCGCTCAGCTCCGCTCCAGGTCCTCC	NotI
<i>h</i> ASC <i>syn</i> 92 STOP rev	AAAAAAGC GGCCGCTCAGCCCTGGTGCGTGGCCGCCTC	NotI

2.1.6. RNA and DNA Oligonucleotides

Table 6 Applied RNA oligonucleotides (syn: synthetic; ivt: in vitro transcribed; s: sense; as: antisense; hp: hairpin; nt: nucleotides; AF: AlexaFluor; RVL: Rabies Virus Leader; FAM: Carboxyfluorescein; DZT: DNAzyme target; RZ: ribozyme)

RNA	company	modification	name	sequence	MW [kDa]
syn	Metabion	5'OH	19 s	GCAUGCGACCUCUGUU UGA	6.2
syn	Metabion	5'OH	19 as	UCAAACAGAGGUCGCA UGC	6.3
syn	Eurogentec	5'PPP	19 s	GCAUGCGACCUCUGUU UGA	6.5
ivt	Ambion (Kit)	5'PPP	18 s/as hp	GGCAUGCGACCUCUGU UUGAUCAAACAGAGGU CGCAUGCC	13.3
ivt	Ambion (Kit)	U-AF488	18 s/as hp	GGCAUGCGACCUCUGU UUGAUCAAACAGAGGU CGCAUGCC	>13.3
syn	IBA	5'AF488	25 s	GCUUGUCGGGAGCGCC ACCCUCUGC	8.8
syn	biomers.net	5'OH	25 as	GCAGAGGGUGGCGCUC CCGACAAGC	8.3
ivt	Ambion (Kit) or recombinant T7 Pol	5'PPP	RVL 58nt	ACGCTTAACAACCAGA TCAAAGAAAAACAGA CATTGTCAATTGCAAA GCAAAAATGT	18.0
syn	Biomers.net	5' 6-FAM	27 s	ACGCUUAACAACCAGA UCAAAGAAAAA	9.3
syn	Biomers.net	-	27 as	UUUUUCUUUGAUCUGG UUGUUAAGCGU	8.5
ivt	Ambion (Kit)	5'PPP	DZT	GGGGAAUU GUGAGCG G	
ivt	Ambion (Kit) or recombinant T7 Pol	-	RZ	NNN GCUAGCCAUGGU CCCAGCCUCCUCGCUG GCGGCUAGUGGGCAAC AUGCUUCGGCAUGGCG AAUGGGACUUUAAAC	24.2

Table 7 Applied DNA oligonucleotides (syn: synthetic; s: sense; as: antisense; hp: hairpin; FAM: Carboxyfluorescein; DZ: DNAzyme)

DNA	company	modification	name	sequence	MW [kDa]
syn	Biomers.net	-	11 hp	GGGCTAGGCGGGCGACC GCCTAGCCC	8.0
syn	Biomers.net	-	18 hp	CACTATAGGGCTAGGCGG GCGACCGCCTAGCCCTAT AGTG	12.3
syn	Biomers.net	-	25 hp	TACGACTCACTATAGGGC TAGGCGGGCGACCGCCTA GCCCTATAGTGAGTCGTA	16.6
syn	Thermo Scientific	3' 6-FAM	11 hp	GGGCTAGGCGGGCGACC GCCTAGCCC	8.6
syn	Thermo Scientific	3' 6-FAM	18 hp	CACTATAGGGCTAGGCGG GCGACCGCCTAGCCCTAT AGTG	12.9
syn	Thermo Scientific	3' 6-FAM	25 hp	TACGACTCACTATAGGGC TAGGCGGGCGACCGCCTA GCCCTATAGTGAGTCGTA	17.2
syn	Biomers.net	-	8 s	GGGCTGGG	2.5
syn	Biomers.net	-	8 as	CCCAGCCC	2.3
syn	Thermo Scientific	5-Br-dC	11 hp	GGGCTAGGCGGGCGACC GCCTAGCCC	8.8
syn	Biomers.net	5-Br-dU	18 hp	CACTATAGGGCTAGGCGG GCGACCGCCTAGCCCTAT AGTG	12.8
syn	Biomers.net	-	35 s	GTGTTGATGAAGGGGGGC TATAAAAGGGGGTGGGG	11.1
syn	Biomers.net	5' ATTO488	35 as	CCCCACCCCTTTTATAG CCCCCTTCATCAACAC	11.1
syn	MWG	-	DZ I	CTCACA GGCTAGCTACAA CGA TTCCCC	8.2
syn	MWG	-	DZ II	TCCGCTCA GGCTAGCTAC AACGA AATTCCCC	9.5

2.2. Methods

2.2.1. Molecular Biological Methods

Molecular biology manipulations were conducted according to standard protocols (Sambrook *et al.* 1989) or as described subsequently.

2.2.1.1. Molecular Cloning

All PCR primers (Table 5) were designed considering melting temperature, overlap with the template DNA, appropriate restriction sites and harbored a 5' overhang of 6 adenosines for improved restriction efficiency. Standard PCR reaction mixes contained 10 µl 2x Phusion[®] Flash High-Fidelity PCR Master Mix (Finnzymes, Espoo, Finland), 8.5 µl water, 0.5 µl template DNA (~30-100 ng/µl in water) and 0.5 µl of each reverse and forward primer (50 pmol/µl in water). PCR cycles were applied as follows:

- 1) 98 °C → 30 s
- 2) 98 °C → 1 s
- 3) 57 °C → 5 s
- 4) 72 °C → 15 s / 1kb
repeat 2) – 4) 30 to 35 times
72 °C → 300 s

Subsequently, the respective restriction enzymes and buffers were added straight to the PCR reactions. Amounts added varied depending on the duration of the digest (~ 3 hours or overnight) or the supplier's instructions for double digests. Typically 2.5 µl 10 x buffers and 1.25 µl of each enzyme were added to the 20 µl PCR mix. Destination vectors were treated accordingly and in addition dephosphorylated by addition of Fast Alkaline Phosphatase for an hour after the digest. Enzymes were inactivated and released from the DNA strands by incubating at 65 °C for 10 min.

PCR products were analyzed on 1 % (w/v) agarose gels (TAE buffer system: 50x TAE – 242g TRIS base, 100 ml 0.5M EDTA pH 8.0, 57.1 ml glacial acetic acid, ad 1 l H₂O) and extracted from excised gel slices with a Metabion (Martinsried, Germany) gel extraction kit. DNA was then eluted in a volume of 30 µl water.

For ligations the digested PCR fragment and vector were added in a 4:1 ration (usually 8 µl of gel extracted insert and 2 µl of vector) with 2 µl 10x ligation buffer and 1 µl T4 DNA ligase in a total volume of 20 µl. Ligation reactions were carried out at 22 °C for 1 hour or

over night at 19 °C and terminated by a 10 min incubation at 65 °C. 10 µl of the ligation mix were then transformed into *E. coli* XL1 blue cells. Utilized vectors are listed in Table 4.

2.2.1.2. Site Directed Mutagenesis

Site directed mutations were introduced, according to the Quikchange protocol from Stratagene (Agilent Technologies, Santa Clara, CA, USA), using mutagenic primers and the Phusion[®] Flash High-Fidelity PCR Master Mix (Finnzymes, Espoo, Finland) analog to the standard PCR protocol but with 0.5 µl of 20 pmol/µl mutagenic primers. Subsequent to the amplification of the whole plasmid in 20 PCR cycles, methylated template plasmid containing the wildtype sequence was digested by addition of DpnI. 10µl of the PCR mix were then transformed into *E. coli* XL1 blue cells. Used mutagenic primers are listed in Table 5.

2.2.1.3. Transformation

Transformations were performed with chemically competent cells. 10 µl from ligation mixtures or 0.5 – 2 µl of purified Plasmid (50 – 300 ng/µl), respectively were added to 70 µl of competent cells and incubated on ice for 15 min. Cells were then heat shocked (45 s at 42 °C) followed by a 2 min incubation on ice, addition of 900 µl LB medium and subsequent incubation on a thermo shaker at 37 °C for 1 hour to establish antibiotic resistance. Bacterial hosts that were used are listed in Table 3.

2.2.1.4. Plasmid Preparation

Plasmids were obtained from 4 ml of overnight culture of *E. coli* XL1 blue cells after lysis and extraction from cell pellets using a Metabion Miniprep kit in a volume of 50 µl water with typical concentrations between 50 and 300 ng/µl.

2.2.1.5. Bacmid Preparation

pFBDM vectors were transformed into *E. coli* DH10 MultiBac cells for integration into bacmids. After heat shock, cells were taken up into 900 µl 2xYT medium and incubated for at least 5 hours shaking at 37°C to establish antibiotic resistance. Cells were then plated on LB agar containing gentamycin, kanamycin, tetracyclin, IPTG and X-Gal for blue white screening, and selection of a colony with an integrated plasmid (white). Bacmids were prepared from a 200 ml overnight culture using a Qiagen Midiprep kit. Ethanol precipitated

bacmid DNA was usually taken up in 100µl water typically yielding bacmid concentrations around 1 µg/µl. Purified bacmid DNA was stored at 4°C.

2.2.2. Protein Biochemical Methods

2.2.2.1. Protein Expression in Insect Cells

To generate a first generation of baculovirus, 5 µg of bacmid DNA were pre-incubated with 2 µl FuGene (Roche, Penzberg, Germany) and 200 µl of High Five medium (Invitrogen, Karlsruhe, Germany) for 45 min at room temperature. Meanwhile 2 ml of High Five insect cells (Invitrogen, Karlsruhe, Germany), freshly diluted in High Five medium to 0.25 Mio/ml cell density, were transferred to a 6 well tissue culture plate and incubated at 27.5 °C. After transfection with the premix, infected cells and one uninfected control per plate were incubated for 48-60 h at 27.5 °C as adhesive culture. The supernatant containing virus generation 0 (V₀) could then be collected and was used for amplification of the viral titer by transferring it to 50 ml freshly diluted 0.5x10⁶ ml High Five insect cells in 500 ml flasks. Cells were incubated for 3-4 days at 27.5 °C and 85 rpm in shaking culture. Cell growth was monitored during this and cells were diluted if necessary to prevent growth over 3x10⁶ cells per ml. Normally, cell growth would be arrested after 1-2 days and the culture is further incubated for 1-2 more days for best expression results. V₁ could be obtained in the supernatant after spinning down the cells and was used for further up-scaling of the expression. Up to 3 virus generations were produced in increasing culture volumes for final expression, which was performed in 1 l cultures in 5 l flasks with according concentrations. Cells were harvested by centrifugation, flash frozen in liquid nitrogen and stored at -20°C. The supernatant containing high viral titers was utilized for further expression and stored at 4 °C.

2.2.2.2. Protein Expression in *E.coli*

Expressed constructs with vectors and expression host are listed in Table 8. Pre-cultures of 30 ml per 3 l expression were inoculated with 1 colony of freshly transformed and plated cells and grown over night at 37 °C in a shaking incubator. 3 l LB were then inoculated with pre cultures 1:100 and grown in shaking culture to an OD₆₀₀ of 0.8 at 37°C. Protein production was started by addition of 0.13 – 0.5 mM IPTG and the cells were shaken over night at 18°C. After harvesting, collected cell pellets were flash frozen in liquid nitrogen and stored at -20 °C.

2.2.2.3. Protein Purification

Fresh or thawed cell pellets were resuspended in lysis buffer (~30 ml/3 l bacterial or 1 l insect cell pellet) and disrupted by sonication (15 min, duty cycle 7, output control 50% for *E. coli* or 5 min and additional stirring of the lysate on ice for insect cells). Whole cell lysate was spun for 30 minutes at 16000 rpm in a Sorvall (Newport Pagnell, UK) centrifuge, SS-34 rotor, prior to draining the supernatant off the pellet of cell debris and keeping it for downstream purification. Supernatants sometimes had to be either filtered or centrifuged again to remove remaining insoluble particles.

Depending on the protein properties and presence of a tag, purification steps included affinity chromatography, ion exchange chromatography (IEC) and size exclusion chromatography (SEC or gelfiltration), typically as last step. All purification steps were followed by analyzing the respective protein containing fractions by SDS PAGE (discontinuous Laemmli-system in TGS buffer: 10x – 720.5g Glycine, 0.5 l 10% w/v SDS, 151.5 g TRIS base, ad 5 l H₂O) (Laemmli 1970). Given a high enough purity fractions from size exclusion chromatography were then pooled and the proteins concentrated *via* ultra filtration in centrifugal filter devices (Millipore, MA, USA) with an appropriate molecular weight cutoff. Depending on their stability and experimental use proteins were typically concentrated to between 1 and 20 mg/ml. Respective purification steps are described subsequently in detail and are listed together with applied buffers for all constructs in Table 8.

2.2.2.3.1. Glutathione-S-Transferase Affinity Chromatography

Glutathione-S-transferase (GST) tagged constructs were in a first purification step applied to a glutathione coupled sepharose resin. Elution of the protein was achieved by competition with buffer containing free glutathione. The GST-tag was cleaved subsequently with PreScission protease (GE Healthcare, München, Germany) followed by either gelfiltration or another GST-affinity chromatography step to retain the free tag.

Table 8 Constructs, Expression host and Purification Steps

Construct	Expression host	Purification step	Buffer
pET28 6xHis - LGP2 RD	<i>E. coli</i> Rosetta DE3	Nickel-NTA	30mM TRIS-HCl pH 7.5; 10/50/100/250mM Imidazole; 0.1M NaCl; 5% Glycerol; 1mM β -Mercaptoethanol
		SP-Sepharose	30mM TRIS-HCl pH 7.5; 0.1/1M NaCl, 5% Glycerol; 2mM DTT
		Superdex S75	30mM TRIS-HCl pH 7.5; 0.15M NaCl, 5% Glycerol; 2mM DTT
pET21a LGP2 RD (537/543/ 549-678)	<i>E. coli</i> Rosetta DE3	SP-Sepharose	30mM TRIS-HCl pH 7.5; 0.1/1M NaCl, 5% Glycerol; 2mM DTT
		Superdex S75	30mM TRIS-HCl pH 7.5; 0.15M NaCl, 5% Glycerol; 2mM DTT
pFBDM 6xHis - LGP2 full	High Five Insect Cells	Nickel-NTA	30mM TRIS-HCl pH 7.5; 10/50/100/250mM Imidazole; 0.1M NaCl; 5% Glycerol; 1mM β -Mercaptoethanol
		Q-Sepharose	30mM TRIS-HCl pH 7.5; 0.1/1M NaCl, 5% Glycerol; 2mM DTT
		Superdex S200	30mM TRIS-HCl pH 7.5; 0.15M NaCl, 5% Glycerol; 2mM DTT
pET21a RIG-I RD (802-925)	<i>E. coli</i> Rosetta DE3	SP-Sepharose	30mM TRIS-HCl pH 7.5; 0.1/1M NaCl, 5% Glycerol; 2mM DTT
		Superdex S75	30mM TRIS-HCl pH 7.5; 0.15M NaCl, 5% Glycerol; 2mM DTT
pET28 (N-His) /pET21a (C-His) 6xHis – AIM2 (various)	<i>E. coli</i> Rosetta DE3	Nickel-NTA	30mM TRIS-HCl pH 7.5; 10/50/100/250mM Imidazole; 0.1M NaCl; 5% Glycerol; 1mM β -Mercaptoethanol
		SP-Sepharose	30mM TRIS-HCl pH 7.5; 0.1/1M NaCl, 5% Glycerol; 2mM DTT
		Superdex S200	30mM TRIS-HCl pH 7.5; 0.15M NaCl, 5% Glycerol; 2mM DTT
pET21a AIM2 full length	<i>E. coli</i> Rosetta DE3	Heparin	30mM TRIS-HCl pH 7.5; 0.4/1M NaCl, 5% Glycerol; 2mM DTT
		Superdex S75	30mM TRIS-HCl pH 7.5; 0.15M NaCl, 5% Glycerol; 2mM DTT
		SP-Sepharose	30mM TRIS-HCl pH 7.5; 0.1/1M NaCl, 5% Glycerol; 2mM DTT
pET21a AIM2 HIN (various)	<i>E. coli</i> Rosetta DE3	Heparin	30mM TRIS-HCl pH 7.5; 0.4M NaCl, 5% Glycerol; 2mM DTT
		SP-Sepharose	30mM TRIS-HCl pH 7.5; 0.1/1M NaCl, 5% Glycerol; 2mM DTT
		Superdex S75	30mM TRIS-HCl pH 7.5; 0.15M NaCl, 5% Glycerol; 2mM DTT
pGEX6P2 GST-ASC (various)	<i>E. coli</i> Rosetta DE3	GSH-Sepharose	50mM HEPES or MES pH 6.5; 0.15M NaCl, 5% Glycerol; 2mM DTT; 0/20mM Glutathione
		Superdex S75	50mM HEPES or MES pH 6.5; 0.15M NaCl, 5% Glycerol; 2mM DTT

2.2.2.3.2. Nickel Affinity Chromatography

Constructs containing an N-terminal 6xHis-Tag were initially purified by immobilized metal affinity chromatography. 2 ml (~4 ml slurry in 20% ethanol) of nickel charged NTA resin (Qiagen) were filled into an empty Econo-Pac column (Biorad, CA, USA) suited for gravity flow chromatography, rinsed with water and then equilibrated with 5 column volumes (cv) of lysis buffer containing 10 mM imidazole to reduce unspecific binding. After loading the protein sample to the resin, unspecifically bound protein was removed by a wash step with 5 cv of equilibration buffer. To ensure high protein purity in the elution fractions two additional washing steps were applied with 2x 2 cv each of buffers containing 50 mM and 100 mM imidazole, respectively. Finally 3-5 x 2 cv of elution buffer containing 250 mM imidazole were applied to the column. Fractions were collected accordingly for all wash and elution steps and further analyzed by SDS PAGE. Depending on the protein purity pooled fractions were either dialyzed in low salt buffer for subsequent ion exchange chromatography or concentrated for size exclusion chromatography.

2.2.2.3.3. Heparin Affinity Chromatography

DNA binding constructs were initially purified with a 5 ml HiTrap Heparin HP column allowing for a higher salt concentration in the lysis / binding buffer than for ion exchange chromatography. Binding buffers usually contained 400 mM NaCl and elution (1.5 ml fractions) of bound proteins was achieved by applying a 20 cv linear gradient from 0 to 100 % 1 M NaCl high salt buffer. Buffers used had physiological pH or slightly lower. Due to insufficient purity this was usually followed by dialysis of the pooled fractions to low salt buffer and cation exchange chromatography.

2.2.2.3.4. Dialysis

In order to change the buffer of a large volume of protein it was subjected to overnight dialysis in a buffer rinsed nitrocellulose tubing (Roth, Karlsruhe, Germany) with stirring, in 2 l of dialysis buffer (usually low salt binding buffer for IEC) at 4°C.

2.2.2.3.5. An- and Cation Exchange Chromatography

Considering their theoretical pI, constructs were either subjected to an- or cation exchange chromatography as a first purification step or following affinity chromatography.

For proteins with pI values above 7.5 a cation exchange 5 ml HiTrap SP Sepharose FF column was used in a buffer system with a pH about 2 units below the pI, if applicable. Proteins with a pI below 7.5 were purified applying an anion exchange 5 ml HiTrap Q Sepharose FF column and a buffer about 2 pH units above the pI. Columns were equilibrated with 5 cv low salt binding buffer prior to loading the proteins with a membrane pump (GE Healthcare), followed by a 5 cv wash step with binding buffer. The binding buffer contained 100 mM NaCl and elution of the proteins was achieved by applying a linear gradient of 20 cv from 0 to 100 % high salt buffer containing 1 M NaCl on an ÄKTA system. Fractions of 1.5 ml were collected. Pooled fractions were then concentrated for size exclusion chromatography.

2.2.2.3.6. Size Exclusion Chromatography (SEC or Gelfiltration)

For preparative SEC either a Superdex S200 26/60 (mainly proteins > 70 kDa) or Superdex S75 26/60 (mainly smaller proteins < 70kDa) have been used with an ÄKTA Purifier. Protein samples were concentrated to at least 2 ml and spun in a tabletop centrifuge (Eppendorf, Hamburg, Germany) for 10 min at 13000 rpm and 4°C before injection to the column. Fractions of 1 ml were collected during elution.

Analytical SEC was carried out on an ÄKTA Ettan system using either Superose 6 or 12 3.2/30 and sample loading volumes of 10 µl.

2.2.3. Crystallographic Methods

2.2.3.1. Crystallization

LGP2 RD was screened for crystallization using common screens from Qiagen and Jena Bioscience. Initial screens were setup in 96 well sitting drop plates with a Hydra II-Plus-One robot (Matrix Technologies, Thermo Fisher Scientific, USA) and drop sizes varied between 0.1 and 0.5 µl. Crystallization screening was tested with and without addition of TCEP. LGP2 RD was crystallized using hanging drop vapor diffusion by mixing 2 µl of protein solution at 13 mg/ml protein concentration with 2 µl of the reservoir solution (500 µl: 1.5 M (NH₄)₂SO₄, 100 mM TRIS-HCl pH 8.5, 0.5 mM TCEP). Crystals grew after several weeks at 21°C.

RIG-I RD/RNA and AIM2/DNA complexes were screened using a Phoenix robot (Art Robbin's, Sunny Vale, CA, USA) and common screens and were refined by either hanging or sitting drop methods.

2.2.3.2. Crystallographic Data Collection of LGP2 RD

Prior to flash freezing in liquid nitrogen, the LGP2 RD crystals were soaked with 2 mM mercury acetate for 10 minutes and then transferred for 5-10 seconds into the reservoir solution additionally containing 2.5 M (NH₄)₂SO₄. The final diffraction data was collected at the theoretical absorption edge of mercury with a wavelength of 1.009 Å at ESRF beamline ID 14-4 (European synchrotron radiation facility, Grenoble, France).

2.2.3.3. Structure Determination of LGP2 RD

2.2.3.3.1. Theoretical Background

Diffraction experiments only yield the intensities of scattered waves arising from the atomic distribution of the molecule within a lattice, while the phases are lost. Hence, in order to produce an interpretable image from experimental data, it is necessary to determine the associated phases. Basically, no formal relationship between amplitudes and phases exist, so they can only be accessed *via* the molecular structure or electron density itself.

Various methods have been established to gain estimates of phases that can then be used to derive more accurate phase values. Such methods include direct methods that require extremely high resolution and completeness, molecular replacement that employs homolog structures with significant sequence identity, and various methods involving heavy metal soaking or derivatization as a means of introducing anomalous scatterers.

The electron density function is given as a Fourier transform of the structure factors that are represented by amplitudes F_{hkl} and phases ϕ_{hkl} :

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \cdot e^{-2\pi i \cdot (hx + ky + lz - \phi_{hkl})}$$

The Patterson function however, replaces the structure factors with the squared amplitudes only, whose values are proportional to the diffraction intensities, leaving phases aside:

$$\rho(u, v, w) = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 \cdot e^{-2\pi i \cdot (hu + kv + lw)}$$

with

$$|F_{hkl}|^2 \propto I_{hkl}$$

Thus, the Patterson Function can be directly calculated from the experimental data obtained in the diffraction experiment and a Patterson map, which represents interatomic distances, can be derived.

For molecular replacement the Patterson maps of a suited search model and the calculated experimental Patterson map are compared and a solution is found when best congruency is reached. In this approach six dimensions have to be considered, that can however be split up in the model search process. First a three dimensional rotational function is applied to determine the relative orientation of the unknown structure and secondly the position is derived by translation in three directions. Ideally the so calculated position of the molecule can be used to derive phases that should be accurate enough to allow for obtaining an electron density map (Taylor 2003).

A newer and more accurate approach for MR uses maximum-likelihood (Read 2001). It assesses the agreement of the model derived theoretical data with the experimental one by using probabilities. The model to be tested includes, aside from the structure, orientation and position of that template in the unit cell of the target, also parameters describing the sizes of different sources of error. Thus, replacement with weak homology models and in cases of many molecules in the asymmetric unit becomes more efficient.

Another means of solving the phase problem is the introduction of anomalous scatterers, i.e. heavy atoms into the protein crystals. This is either achieved by soaking the crystals in heavy metal solutions or by growing selenomethionine containing protein crystals. The simplest and least radiation damage causing phasing methods using anomalous scattering are single-wavelength anomalous dispersion (SAD) or single isomorphous replacement with anomalous scattering (SIRAS). Thereby, the exposure of the heavy atom derivative crystals to a certain wavelength x-ray beam that ideally lies at the absorption edge of the included heavy atoms, for example 1.009Å for mercury, causes anomalous dispersion. In addition to elastic scattering the incident x-ray wave is absorbed, causing a slight retardation and phase shift of the scattered wave. Even though, resulting phases still underlie a twofold ambiguity, sufficiently good phase estimates can be gained by including probabilities and use of density modification methods (Dodson 2003; Taylor 2003; Taylor 2010).

2.2.3.3.2. *Solution of the LGP2 RD Structure*

Diffraction data were processed with XDS (Kabsch 1993). Heavy atom sites were located using autoSHARP (Vonrhein *et al.* 2007) and an initial experimental electron density map could be derived. Due to high sequence similarity, the structure was then however

determined by molecular replacement with Phaser (McCoy *et al.* 2007a), using human RIG-I RD (PDB entry 2QFB,(Cui *et al.* 2008)) as a search model. The resulting $2F_o-F_c$ and F_o-F_c electron density maps allowed initial manual model building. The model was iteratively refined by cycles of bulk solvent correction, overall anisotropic B factor refinement, positional refinement and overall B factor refinement with CNS 1.2 (Brunger 2007) and Phenix (Adams *et al.* 2002). Initial non crystallographic symmetry restraints were gradually removed in the final cycles of the refinement, to allow some structural variations. Manual model building was performed with Coot (Emsley *et al.* 2004) and solvent atoms added with CNS 1.2.

2.2.4. RNA and DNA Biochemistry

Double strands were annealed in a thermo cycler by first heating to 95°C for 5 min, then stepwise lowering of the temperature to 4°C and stored at -20°C. 3' ends were unmodified, if not stated differently. Fluorescent labels were either located at 5' or 3' ends or throughout the oligonucleotide when incorporated by *in vitro* transcription. RNA and DNA hairpins were annealed by heating at 95°C for 5 min and immediate incubation on ice to avoid annealing of two complementary strands rather one intrinsically. DNA oligonucleotides used for AIM2 binding assays and co-crystallization are listed in Table 7.

2.2.4.1. RNA Preparation

In general RNA was handled according to standard procedures with transcription and T7 Polymerase purification protocols used as described in the “Handbook of RNA Biochemistry” (Hartmann *et al.* 2005).

5'-triphosphate containing RNA for binding assays was either obtained by *in vitro* transcription using the Ambion (Austin, TX, USA) MEGAscript™ Kit (or purchased from Eurogentech (Köln, Germany)). RNAs from *in vitro* transcription mixes were purified using MicroSpin™ G-50 columns (GE Healthcare) or reversed phase chromatography (μ RPC C2/C18 2.1/100 column; solvents: 8.6 mM TEA, 100 mM HFIP in either H₂O for loading or methanol for elution) and checked for purity by denaturing urea PAGE (12-20 % (w/v) acryl-/bisacrylamide, 6M urea, 1x TBE). RNAs were fluorescence labeled by either incorporating AlexaFluor 488-5-UTP during *in vitro* transcription or purchased from IBA with AlexaFluor 488 as 5' modification. Unmodified oligonucleotides were purchased from Metabion or biomers.net. Used RNAs are listed in Table 6. RNAs for co-crystallization with RIG-I RD

were also obtained by *in vitro* transcription using either T7 RNA polymerase purified in the lab or the Ambion MEGashortscript™ Kit. Dideoxy-UTP or AlexaFluor 488-5-UTP were thereby incorporated to achieve homogenic 3'-ends by polymerization break-off.

2.2.4.2. Ribozymes and DNAzymes

As another means of producing homogenic RNAs, ribozyme transcription and auto-cleavage was conducted. Ribozyme templates used were derived from a human delta virus (HDV) ribozyme containing vector (Walker *et al.* 2003). The ribozyme template containing vector was first amplified, then linearized by XhoI digest and then purified by gelextraction. After a 37 °C over night *in vitro* transcription step the reaction mixture was subjected to a thermo cycle optimized for ribozyme cleavage (6x repeat: 1 min at 72 °C, 5 min at 65 °C, 10 min at 37°). Ribozyme cleavage products were analyzed in 6 M urea/ 12 % polyacrylamide (in 1x TBE) gels and detected by UV shadowing.

DNAzymes were also tested under standard conditions to gain homogeneous 3' ends and short RNAs (Schubert *et al.* 2003). Applied oligonucleotides are listed in Tables 6 and 7.

2.2.5. Biochemical Assays

2.2.5.1. Fluorescence Anisotropy Measurements

Fluorescence anisotropy experiments were performed with a FluoroMax-3P fluorimeter (HORIBA Jobin Yvon, Unterhaching, Germany), equipped with Glan-Thompson prism polarizers and a temperature stabilized cuvette holder (connected to a Haake F3 thermostat). Typically, 1 ml of buffer (30 mM TRIS-HCl pH 7.5, 150 mM NaCl, 2 mM DTT, and 10 µM ZnCl₂) and 37 nM RNA (5' AlexaFluor 488 labeled/ 5'OH dsRNA, 25 bp) were pre-equilibrated in a quartz cuvette at 12 °C. Protein samples were added in a stepwise manner. After mixing and 4 minutes of incubation to reach equilibrium, anisotropy data was collected using an excitation wavelength of 495 nm and monitoring the emission at 516 nm. The band pass was 5 nm for excitation and 5 nm for emission. A maximum number of ten trials were performed until minimal deviation of the signal was reached. For the competition assays 800 µl of buffer, 40 nM of an *in vitro* transcribed hairpin RNA (AlexaFluor 488-5-UTP incorporated), and 470 nM protein were pre-equilibrated in a quartz cuvette at 12°C. Unlabeled RNA species were added in a stepwise manner and the drop in anisotropy was monitored after mixing and 4 minutes of re-equilibration.

Data were fitted applying a simplified single-site binding model by non-linear least square fitting using the Origin (Northampton, MA, USA) data analysis software:

$$\Delta A = \frac{A_{max} * x}{K_d + x}$$

where ΔA is the measured anisotropy difference, x the applied protein concentration and K_d the deduced dissociation constant.

2.2.5.2. Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays were carried out using 10% (w/v) polyacrylamide native gels in TRIS borate buffer (5x – 54 g/l TRIS, 27.5 g/l borate). Samples contained 5% glycerol, 60 nM 5' AlexaFluor 488 labeled/5'-OH dsRNA (25 bp) and different concentrations of protein, diluted from 5 μ M stock solutions (in 30 mM TRIS-HCl pH 7.5, 150 mM NaCl, 2 mM DTT and 10 μ M ZnCl₂). Samples were incubated for 20 minutes at 4°C prior to electrophoretic analysis. Gels were analyzed with a Typhoon™ scanner (GE Healthcare).

2.2.5.3. Pulldown Assays

To determine physical interaction between C- or N-terminally Hexa-His-tagged *h/mAIM2* and GST-tagged *h/mASC* or its PYD respectively, pulldowns were performed on GSH-Sepharose resin, immobilizing the ASC component. MES buffers at pH 6.5 containing 150 mM NaCl and 2 mM β -mercaptoethanol were used and supplemented with imidazole, more salt (wash step of Ni-NTA immobilized AIM2) or GSH as required. Proposed binding partners were either expressed together (different but compatible vectors, N-terminal 6xHis AIM2) or separately (C-terminal 6xHis-tag for AIM2) in 100 ml cultures of *E.coli* Rosetta (DE3) cells. Constructs were either purified by Ni²⁺- or GSH-affinity chromatography separately. And eluted AIM2 constructs were then loaded to ASC or ASC-PYD immobilized on GSH-Sepharose resin. In addition, cell pellets containing either of the binding partners were mixed prior to cell lysis for co-purification by GSH-affinity chromatography only. Samples from co-expressions were also only purified by GSH-affinity chromatography. Purification steps were applied as described earlier (see chapter 2.1.8.3). Flow through, wash, loaded bead and elution fractions were analyzed by SDS PAGE and for better clarity immunostained after western blotting.

2.2.5.4. Western Blots and Immunostaining

Protein samples from SDS PAGE gels were blotted to Nitrocellulose membranes (Roth, Karlsruhe, Germany) using a Blot® SD Semi-Dry Electrophoretic Transfer Cell from Biorad (Hercules, CA, USA) with transfer buffer containing 25 mM TRIS pH 8.6, 192 mM glycine, 0.05 (w/v) % SDS, 20 % methanol at pH 9.2.

For immunostaining, blots were rinsed several times in TBS-T (10mM TRIS pH 8.0, 150mM NaCl, 0.01% Tween 20). Blots were then blocked against unspecific binding with a 10 % (w/v) milk powder solution in TBS-T for one hour under continuous shaking and subsequently rinsed again. Membranes were probed by applying either a Horseradish Peroxidase (HRP) – coupled primary GST-antibody (GE Healthcare, 1:5000 in 2 % (w/v) milk powder solution in TBS-T) or both, a primary mouse His-antibody (1:2500) and subsequently a secondary HRP-coupled goat anti-mouse antibody (GE Healthcare, 1:5000) for one to two hours.

Accordingly tagged protein was detected by adding the SuperSignal West Pico Chemiluminescent Substrate from Thermo Scientific (Bonn, Germany), subsequent exposure of the probed blots to a Hyperfilm™ ECL™ (GE Healthcare) and film development.

2.2.6. Bioinformatic Methods

2.2.6.1. Sequence Alignments

Multiple sequence alignments were built with ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/>) (Thompson *et al.* 2002) and visualized using ESPript (<http://esprict.ibcp.fr/ESPript/ESPript/>).

2.2.6.2. Calculation of Protein Parameters

Physical and chemical parameters of the recombinant proteins such as molecular weight, theoretical isoelectric point (pI) and extinction coefficients were calculated with ProtParam (Wilkins *et al.* 1999) from the ExPASy Proteomics Server (www.expasy.org/tools/protparam.html).

2.2.6.3. Structure Visualization and Analysis

Structural visualization was achieved with PyMol (<http://www.pymol.org>) and electrostatic surfaces were calculated using the APBS plugin for PyMol (Baker *et al.* 2001).

Structures were superimposed for comparison using DaliLite (Holm *et al.* 2000). Structural conservation of single residues was assessed and visualized by ConSurf (Landau *et al.* 2005).

2.2.6.4. Protein Profile Search

Protein domains, patterns and profiles were identified using InterProScan at EBI (<http://www.ebi.ac.uk/Tools/InterProScan/>) and according structures were available through the protein data bank (PDB, <http://www.pdb.org>) of the Research Collaboratory for Structural Bioinformatics (RCSB).

2.2.6.5. Structural Homology Modeling

Comparative structural homology modeling was performed with MODELLER (Eswar *et al.* 2008) using the Bioinformatic Toolkit accessible online at <http://toolkit.lmb.uni-muenchen.de> (Biegert *et al.* 2006).

2.2.6.6. Secondary Structure Predictions

RNA and DNA secondary structures were predicted with Mfold (<http://mfold.bioinfo.rpi.edu/>). Protein secondary structures were predicted based on sequence alignments with JPred (<http://www.compbio.dundee.ac.uk/www-jpred/>).

2.2.7. Analytical Methods

2.2.7.1. Mass Spectrometry

Mass spectrometrical analysis of AIM2 degradation products was performed by the central protein analysis unit (ZfP) of the Ludwig-Maximilians-University München. Protein identification by MALDI peptide mass fingerprints was employed.

2.2.7.2. Edman-Sequencing

N-terminal degradation of AIM2 was analyzed by Edman-Sequencing at the protein analysis department of Prof. F. Lottspeich at the MPI of biochemistry (Martinsried, Germany) using standard methods.

3. LGP2 – Results

3.1. Full Length LGP2

Full length (FL) LGP2 constructs harboring different tags were entirely insoluble when expressed in *E. coli* and so were helicase domain constructs. LGP2 could however be expressed in *High Five* insect cells. Still, only small amounts of soluble protein were obtained due to partial aggregation. Even in the presence of high salt concentrations during cell lysis the major fraction of LGP2 remained insoluble. Furthermore, purified LGP2 was rather prone to degradation (Fig. 9 A). Also preparations of LGP2 which has a pI around 7, using different buffers (pH 6, 7.5, 8.5) could neither improve protein yield nor stability. Purified LGP2 could not be highly concentrated (~ 1 mg/ml maximum), making crystallization setups impossible.

Protein that had been stored at -80 °C as well as freshly prepared samples failed to exhibit any significant ATPase activity. Also an expected LGP2 dsRNA (25 bp or 40 bp) complex was not able to be detected by analytical gelfiltration. This might however be due to the protein's instability. On the other hand electrophoretic mobility shift assays (EMSAs) (Fig. 9 B) could verify LGP2's ability to bind double stranded RNA as previously hypothesized and as demonstrated for its regulatory domain (See chapter 3.2.). LGP2 is clearly retarding a 27 bp dsRNA in an EMSA, however the affinity of LGP2 to the labeled RNA seems to be low. Even with an approximately 17 times higher concentration of protein, there is still free RNA. An explanation for the weak binding, apart from mentioned problems with LGP2 samples, might be sterical hindrance of binding caused by the fluorescence label on the 5' end of the dsRNA.

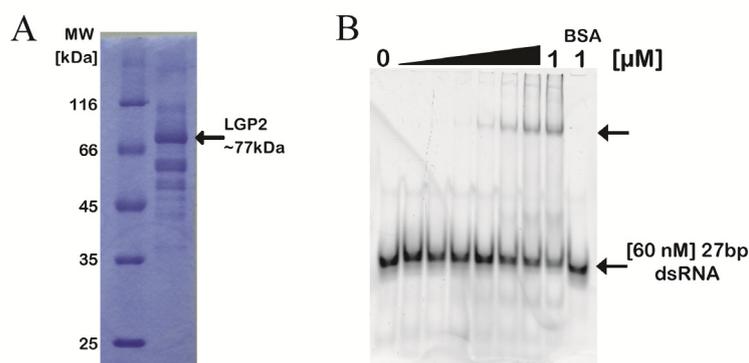


Figure 9 (A) Purified LGP2 sample (with N-terminal 6xHis-tag) from insect cell expression, analyzed on a 15 % SDS PAGE gel. (B) EMSA of a 27 bp Atto488 labeled dsRNA in presence of increasing concentrations of LGP2 analyzed on a 10 % native gel in a TB buffer system.

3.2. The Regulatory Domain of LGP2

Full length LGP2 and helicase domain constructs proved to be too instable or insoluble and were not pursued further. The regulatory domain of LGP2 however emerged as an interesting target. The RD of RIG-I had been shown to be crucial for its nucleic acid sensing and signaling ability. It could be proposed that this might be also the case for LGP2 RD. Hence, gaining detailed knowledge of the molecular properties of LGP2 RD was supposed to help to explain the controversial regulatory effects LGP2 exhibits towards RIG-I and MDA5 signaling.

3.2.1. Constructs and Purification

Constructs for the LGP2 RD were chosen according to sequence alignments with its homologs RIG-I and MDA5 (Figure 10), secondary structure predictions and the structure of RIG-I RD as guide for the domain boundary (PDB: 2QFD, 2QFB) (Cui *et al.* 2008). Constructs containing residues 543 - 678 (native C-terminus), 544 - 671, 549 - 678 or 537 - 678 were expressed with different tags or without any tag in *E. coli* Rosetta cells and behaved similarly during purification. Due to its well established expression and purification protocol and highest solubility only the RD 537 - 678 construct expressed without tag was used for final crystallization setups, and the assays and mutational studies shown. Protein was purified from *E. coli* cell lysate using cation exchange and size exclusion chromatography. LGP2 RD was obtained in high purity but was prone to aggregation when concentrated higher than 10 mg/ml. As an exception a concentration of 13 mg/ml could be reached for the construct containing residues 537 - 678 of LGP2 which was successfully crystallized.

Point mutants of LGP2 RD 537-678 were obtained through site directed mutagenesis and were treated and behaved analogously to the wildtype construct. Only a mutation in the highly conserved Zn^{2+} coordination site, mutant C^{615A}, resulted in an insoluble RD, likely due to compromised protein integrity. All mutated residues are listed and shown in Figures 10 and 16.

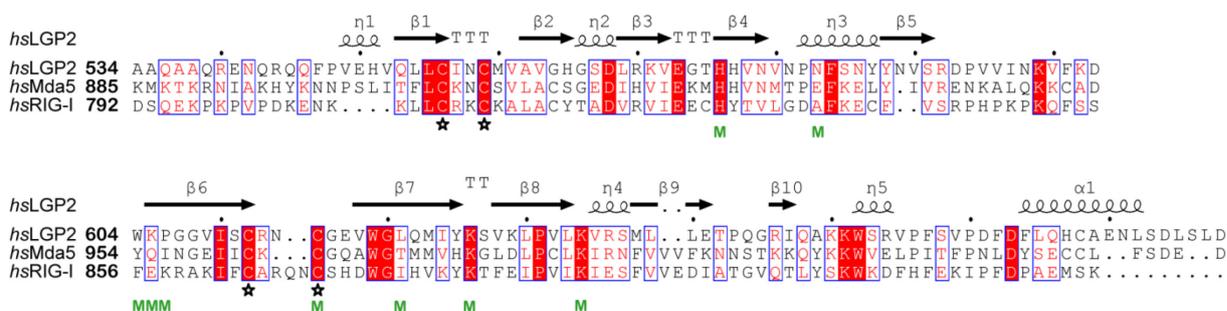


Figure 10 Sequence alignment of RLR RDs with secondary structure annotation according to the LGP2 RD crystal structure. The four invariant Cys residues required for Zn^{2+} coordination are marked with asterisks. Identical residues are depicted in white and shaded in red, homologous residues are shown in red. Point mutated residues are marked with an **M**.

3.2.2. Crystallization and Structure Determination of LGP2 RD

Various LGP2 RD constructs were initially screened for crystallization using different screens, drop sizes and with and without addition of TCEP. However no crystals were obtained for most constructs. The condition that finally led to successful crystallization of LGP2 RD comprising residues 537 – 678 was derived from condition B2 of Jena Bioscience Classic screen II (2 M ammonium sulfate and 100 mM TRIS-HCl pH8.5). Initially, setups under this condition only resulted in crystalline precipitate in the screen and most refinements. Macrocrystals were grown by hanging drop vapor diffusion, with a 500 μ l reservoir (1.5 M $(NH_4)_2SO_4$, 100 mM TRIS pH8.5) and a drop size of 2 μ l protein and 2 μ l reservoir solution, after several weeks of incubation at 21°C. Crystals grew as clusters from the drop boundary but could be separated from each other (Fig. 11 A). Prior to cryo-protectant soaking and flash freezing in liquid nitrogen the crystals were soaked with mercury acetate in case molecular replacement would not be applicable and anomalous scattering would be required to allow for phasing. The best cryo-protection was achieved when soaking crystals in 2.5 M $(NH_4)_2SO_4$ and 100 mM TRIS-HCl pH 8.5. The final diffraction data was collected at the theoretical mercury absorption edge at a wavelength of 1.009 Å at ESRF beamline ID 14-4 (European synchrotron radiation facility, Grenoble, France).

LGP2 RD crystallized in space group $P2_12_12_1$ with four molecules per asymmetric unit. The crystals diffracted to 2.6 Å (Fig. 11 B) and an initial experimental electron density map revealing the localization of four mercury atoms could be obtained employing autoSHARP (Vonrhein *et al.* 2007) (Fig. 11 C). The actual structure could however be solved by molecular replacement with PHASER (McCoy *et al.* 2007b) using CHAINSAW (Stein 2008) adapted coordinates of RIG-I RD as search model. The four copies were initially refined with non-

crystallographic symmetry restraints that were removed during final cycles of refinement. Parts of the experimental $2F_o-F_c$ electron density map and the final refined $2F_o-F_c$ electron density map after molecular replacement are shown in Figure 11 C and D. The obtained model spans residues 544 to 671, with the loop region between residues 593 and 601 not visible in the electron density map. The structure was refined to an R value of 22.7 (free R 28.0) using Phenix and CNS (Adams *et al.* 2004; Brunger 2007) and its geometry is within normal parameters. Crystallographic data and statistics are summarized in Table 9.

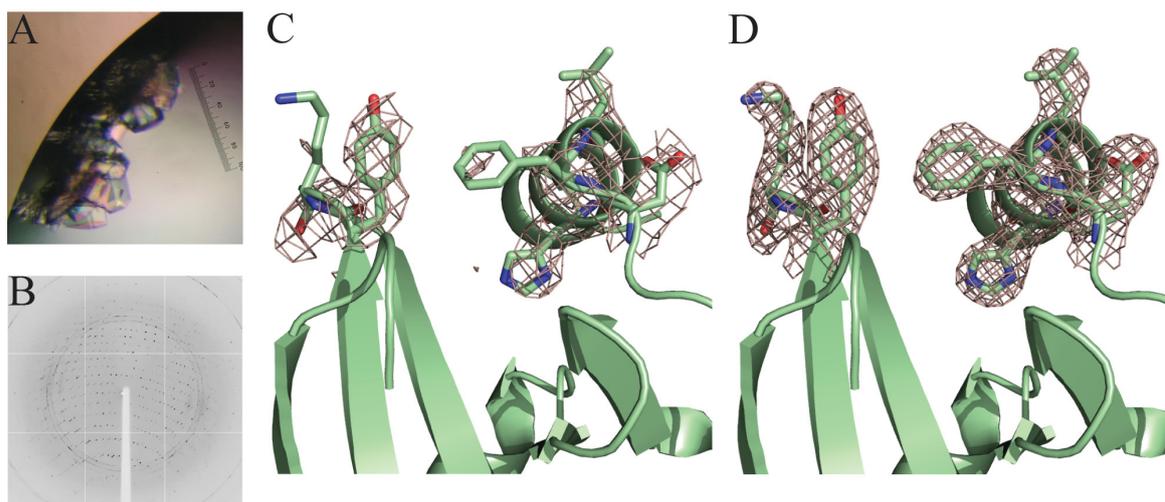


Figure 11 (A) Crystals of LGP2 RD comprising residues 537 - 678. (B) Diffraction pattern of LGP2 RD crystals to 2.6 Å. (C) Initial experimental $2F_o-F_c$ electron density map of LGP2 RD obtained by phasing using anomalous scattering of incorporated Hg^{2+} with autoSHARP. (D) Refined $2F_o-F_c$ electron density map of LGP2 RD acquired by molecular replacement with RIG-I RD. Electron densities are shown at a contour level of 1σ and only around some of the residues for better clarity.

3.2.3. Overall Structure

The regulatory domain of LGP2 is a globular, slightly flattened domain with a concave and convex side and dimensions of approximately 45 Å x 35 Å x 30 Å (Fig. 12 B). It is organized in three leaves, consisting of two four-stranded ($\beta_1, \beta_2, \beta_9, \beta_{10}$ and $\beta_5, \beta_6, \beta_7, \beta_8$) and one two-stranded (β_3, β_4) antiparallel β -sheet. Small 3_{10} helical turns ($\eta_1 - \eta_5$) connect the three β -sheets. The C-terminus contains a short α -helix (α_1). The loop connecting β_5 and β_6 is partially undefined in the crystal structure of LGP2 RD. This and the high B factors apparent for residues at the base of the loop indicate increased flexibility in this region.

Table 9 Crystallographic Data and Refinement Statistics

Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
a, b, c (Å)	a=63.67 b=75.63 c= 147.87
α, β, γ (°)	90, 90, 90
Wavelength (Å)	1.009
Resolution (Å)	50 – 2.6 (2.75 – 2.6) ^a
Reflections _{observed}	105887
Reflections _{unique}	41141
R _{sym}	5.6 (29.3)
I / σ I	13.68 (3.58)
Completeness (%)	92.8 (85.7)
Refinement	
Resolution (Å)	48.0 – 2.6
No. reflections	22420 / 1088
R _{work} / R _{free}	22.7 / 28.0
No. atoms	
Protein	3854
Hg ²⁺	4
SO ₄ ²⁻	5
Water	248
B-factors	
Protein	36.1
R.m.s deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.26
Ramachandran statistics (%)	
most favored	89.4
additionally allowed	9.9
generously allowed	0.7

^aHighest resolution shell is shown in parentheses.

The four-stranded β -sheets are laterally connected by two protruding loops, each containing two highly conserved cysteine residues (C⁵⁵⁶ and C⁵⁵⁹; C⁶¹² and C⁶¹⁵). Together, the four thiol groups of these cysteines coordinate a mercury ion in each of the four molecules in the asymmetric unit (Fig. 12 A, B). Based on the RIG-I RD structure and in accordance with the stability of LGP2 RD in ZnCl₂ supplemented buffers, Zn²⁺ can be assumed to be the physiologically coordinated metal ion in LGP2 RD. It has however been replaced by Hg²⁺ in the structure due to the exposure of the crystals to a high Hg²⁺ concentration.

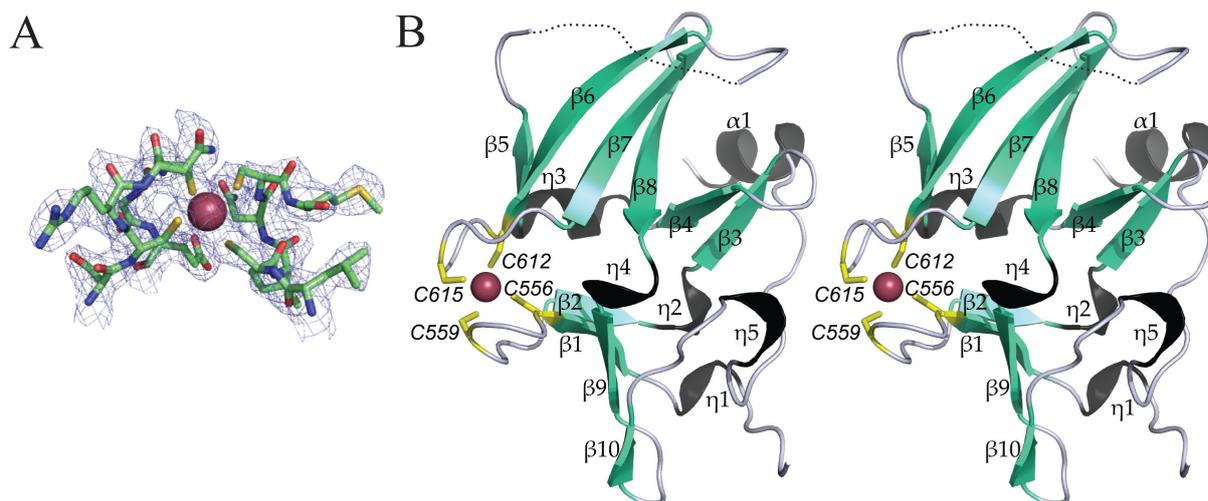


Figure 12 (A) Shows the Zn²⁺ (respectively Hg²⁺) coordination by a 4 Cys cluster in LGP2 RD with the refined 2F_o-F_c electron density map at 1σ. A stereo cartoon representation of the tertiary crystal structure of LGP2 is depicted in (B).

The concave part of the LGP2 RD is highly positively charged implicating the presence of an RNA binding site similar to RIG-I RD.

Two sulfate ions were found to be coordinated on the convex site of the molecule. Even though the RNA binding site is most likely to be located on the opposite, concave region of LGP2 RD these sulfate ions might resemble a second binding moiety with the sulfates mimicking RNA backbone phosphates (Fig. 13). However there is no experimental evidence for this and mutations in this region (N^{853D}) did not alter RNA interaction (see chapter 3.2.6).

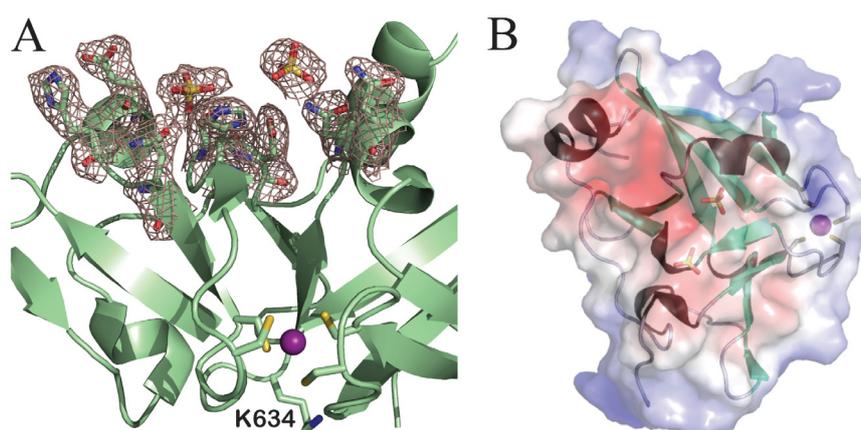


Figure 13 (A) Refined 2F_o-F_c electron density map representation of the two sulfate molecules found coordinated in the LGP2 RD crystal structure and neighboring amino acid side chains. Opposite this coordination site K⁶³⁴ that is located in the center of the positively charged cleft is highlighted. (B) Location of the two sulfate ions in a little conserved region opposite the positively charged proposed RNA binding cleft.

3.2.4. Comparison of LGP2 RD to RIG-I and MDA5 RDs

With the exception of some differences in loop regions that connect secondary structure, the overall fold of LGP2 RD is highly related to that of RIG-I and MDA5 RDs (Fig. 14 A). In particular, the backbone geometry of the metal ion coordinating cluster is highly conserved between RIG-I, MDA5 and LGP2 (Fig. 14 A). The metal in LGP2 is coordinated in a tetrahedral manner by the sulfur atoms of the cysteine cluster C⁵⁵⁶, C⁵⁵⁹, C⁶¹² and C⁶¹⁵. Although a mercury ion is found in this crystal form the geometry should be similar in the presence of Zn²⁺, since crystal structures of RIG-I RD in the presence of mercury and zinc ions have a virtually identical conformation (Cui *et al.* 2008).

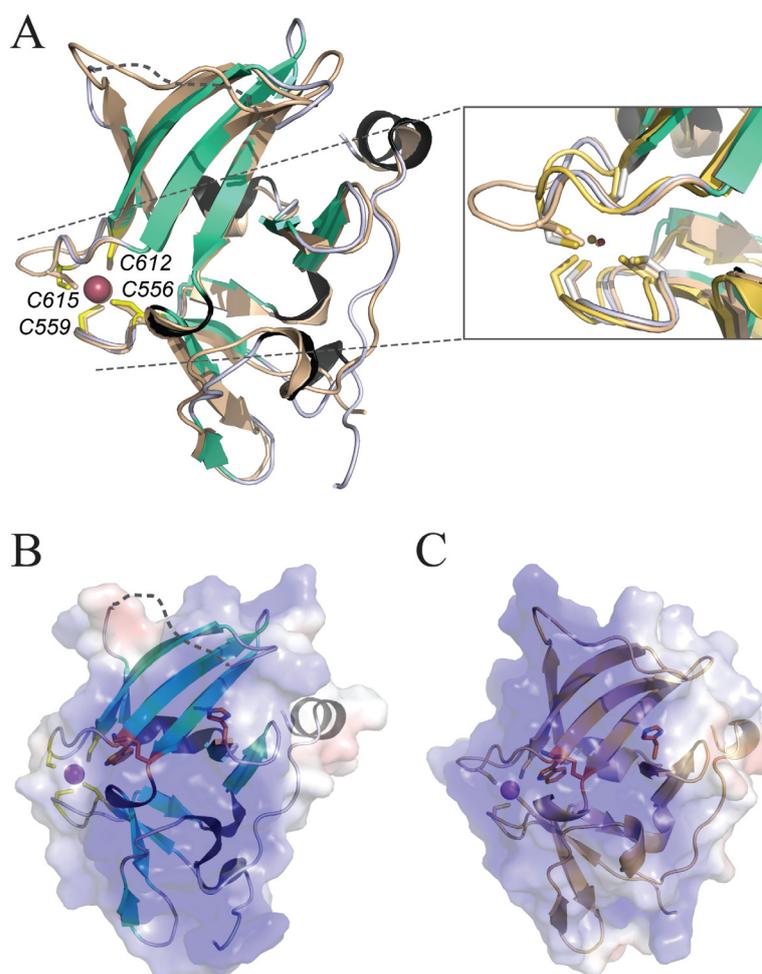


Figure 14 (A) Superposition of the entire RD of LGP2 (green, black, grey) and RIG-I (pale orange) with the Cys cluster and putative RNA binding sites highlighted. Inset: close up view of the superposition of the Zn²⁺ (Hg²⁺)-coordination sites of all three RLR RDs (RIG-I: pale orange, MDA5: yellow, LGP2: grey loops/green sheets). (B+C) Positively charged putative RNA binding clefts of LGP2 (B) and RIG-I RD (C). Conserved W^{619/873} that bounds the 4-Cys cluster towards the domain core, L^{634/888} and H^{576/830} residues are shown in stick representation. Electrostatic surface charge potential ranges from -5 kT (red) to 5 kT (blue).

Tryptophan^{619LGP2} – conserved between LGP2, RIG-I and MDA5 – bounds the Cys-cluster towards the core of the domain (Fig. 14 B, C). Thus, correct formation of the metal coordination sphere is likely to be essential for the fold and integrity of RDs. In support of this, point mutation of the metal coordinating C^{615LGP2} to an alanine resulted in an unstable LGP2 RD, which was insolubly expressed in *E. coli*. Also the severe effect of cysteine point mutations in the zinc-binding cluster on signaling activity of RIG-I (Cui *et al.* 2008) and LGP2 (Pippig *et al.* 2009) correlates with this.

A notable difference in the metal binding site between RIG-I, MDA5 and LGP2 RD is the loop, which connects C^{612LGP2} and C^{615LGP2}. In RIG-I this loop is two residues longer than in LGP2, forming a short β -turn. In contrast, for MDA5 this loop is one amino acid shorter than in LGP2. This indicates that this region could account for functional differences between RIG-I like helicases (Fig. 14 A).

The regulatory domain of RIG-I was shown to specifically bind to RNA with 5'-triphosphates and is suggested to be a main sensor site for 5'-triphosphate containing viral RNA. Key residues important for RIG-I RNA binding were mapped to a groove formed at the interface of the four- and two-stranded β -sheets. These include several lysines as well as a histidine, residues which are well suited to bind to the RNA backbone and the 5'-triphosphate moiety. Intriguingly, two of these residues, K^{888RIG-I} and H^{830RIG-I} are conserved in LGP2 (K^{634LGP2} and H^{576LGP2}, respectively), indicating that this region could also be an important functional site of LGP2 (Fig. 14 B, C).

Furthermore, in LGP2 RD, the groove carries a similarly pronounced positive electrostatic potential that would be ideal for RNA backbone interaction (Fig. 14 B, C). Additionally, this area is flanked by two conserved 3_{10} turns (η 4 and η 5), which are possible phosphate recognition sites (Fig. 12 B).

The positively charged cleft of RLR RDs exhibits much higher sequence conservation amongst homologs than the rest of the molecule (Fig. 15). Thus, this area is very likely a common RNA binding site in RLRS, whereas the unconserved site of the molecule could confer the unique regulatory mode of action of LGP2.

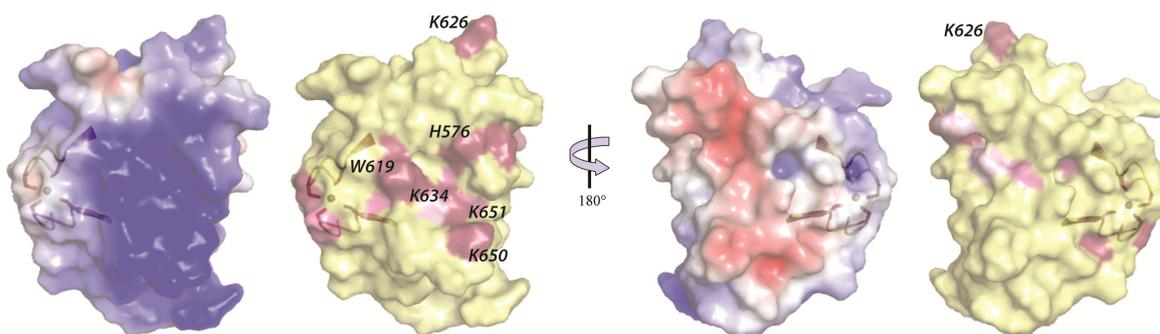


Figure 15 Single residue conservation in RLR RDs is shown in the LGP2 structure with a representation of the surface charge potential (-5 kT red to 5 kT blue) in comparison to a surface representation with invariant residues depicted as purple areas. The conserved Cys cluster is shown in cartoon representation for orientation.

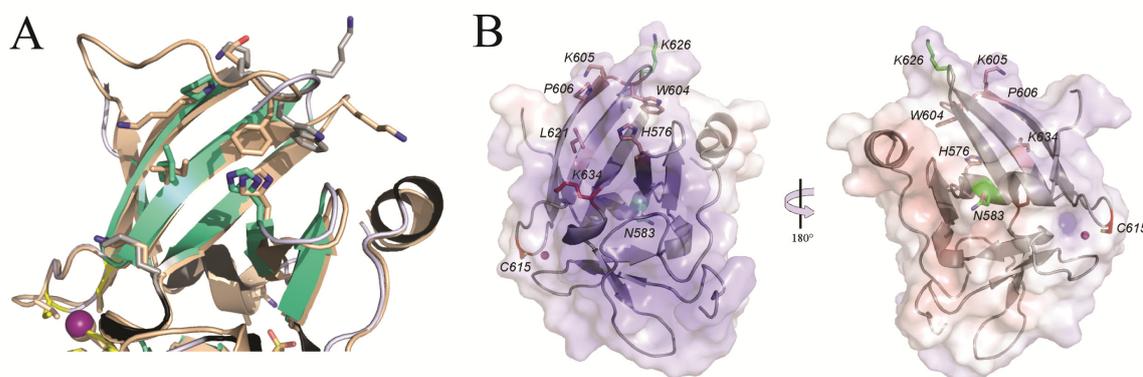


Figure 16 (A) Comparison of the unconserved area of residues around K⁸⁵⁸RIG-I, found to be essential for 5'-triphosphate RNA interaction in RIG-I, by superposition of RIG-I (pale orange) and LGP2 RD (green, black, grey) structures. (B) Point mutated residues for RNA binding studies and their location in LGP2 RD (H⁵⁷⁶Y, N⁵⁸³D, W⁶⁰⁴A, K⁶⁰⁵E, P⁶⁰⁶K, C⁶¹⁵A, L⁶²¹A, K⁶²⁶E and K⁶³⁴E). Electrostatic surface potential is shown for orientation and ranges from -8 kT (red) to 8 kT (blue).

Despite the high similarity of RIG-I and LGP2 RD at initial inspection, some important structural differences arise. RIG-I's K⁸⁵⁸ has for example been shown to be essential for 5'-triphosphate RNA interactions. However this residue is not conserved in LGP2 and is replaced by a proline in this position (P⁶⁰⁶, Fig. 16).

Interestingly, some of these crucial but variable residues are directly adjacent to the flexible loop region between β -sheets 5 and 6. This loop seems to shield the variable residues located in β 6 towards the surface. Hence, LGP2 is not entirely structurally conserved concerning residues shown important for RIG-I activity and RNA affinity. It is therefore

probable that while the general RNA binding area is the same amongst RD domains of the RIG-I like family, minor differences in certain residues confer different specificities towards distinct RNA types.

Based on this, the region around LGP2 P⁶⁰⁶ with the adjacent loop between β 5 and β 6 (Fig. 12, 16) was proposed to confer specificity in RNA binding. Residues to be mutated in order to test this hypothesis were chosen according to the electrostatic surface potential of LGP2 RD and based on the RIG-I RD structure, as well as sequence alignments and previously conducted mutant studies of RIG-I.

Point mutations were introduced for residues homolog to those in positions found crucial for RNA interaction of RIG-I RD (H⁸³⁰ \rightarrow H^{576Y}, K⁸⁵⁸ \rightarrow P^{606K}, I⁸⁷⁵ \rightarrow L^{621A} and K⁸⁸⁸ \rightarrow K^{634E}). Further point mutants were produced for the immediate neighbors of P⁶⁰⁶ (F⁸⁵⁶ \rightarrow W^{604A}, E⁸⁵⁷ \rightarrow K^{605E}). Another mutation was introduced outside the positively charged cleft in proximity to the identified SO₄²⁻ ions, A⁸³⁷ \rightarrow N^{583D}, to examine whether this area is functionally important. Furthermore, a conserved Lys outside the putative RNA binding cleft K⁸⁸⁰ \rightarrow K^{626E} and a Cys in the Zn²⁺ binding site, to prove the cluster's importance for protein integrity and function among the RLR group (C⁸⁶⁹ \rightarrow C^{615A}), were mutated. Residues were either exchanged to uncharged alanines or to inversely charged amino acids, like Glu for Lys. P⁶⁰⁶ was mutated to a Lys, which is found in RIG-I RD in this position.

3.2.5. LGP2 RD Binds to dsRNA in a 5'-Triphosphate Independent Manner

To learn more about the functional sites of LGP2 RD, its binding to different RNA structures was examined. The physiological ligand for LGP2 is unclear; however it has been shown to prefer to bind dsRNA over ssRNA (Murali *et al.* 2008). Its negative regulation of RIG-I dependent signaling *in vivo* may result from competition for viral RNA and thus could be directed against two RIG-I PAMPs, 5'-triphosphate RNA and dsRNA.

According to the findings for RIG-I RD it seemed likely that RD of LGP2 also represents an RNA recognizing element. To confirm this hypothesis RNA oligonucleotides were tested for binding to LGP2 RD in equilibrium binding experiments using fluorescence polarization anisotropy measurements. LGP2 RD was found to bind to a 25mer dsRNA with high affinity and an apparent K_d of 68 \pm 6 nM (see also chapter 3.2.6 and Fig. 18).

Furthermore binding to other RNA species was analyzed by means of competition experiments (Fig. 17). Thereby, a fluorescently labeled dsRNA hairpin was titrated away from LGP2 RD by increasing amounts of unlabelled ssRNA or dsRNA ligands bearing or lacking 5'-triphosphates. 5'-OH ssRNA, corresponding in length and sequence to the labeled probe, is a poor competitor, indicating that LGP2 RD does not efficiently bind to ssRNA. On the other hand, 5'-OH dsRNA with blunt ends is an efficient competitor. Since dsRNA with two blunt ends binds more efficiently than a 5'-triphosphate hairpin with a similarly long stem region, it is likely that certain RNA end structures could enhance specific dsRNA binding by LGP2 RD.

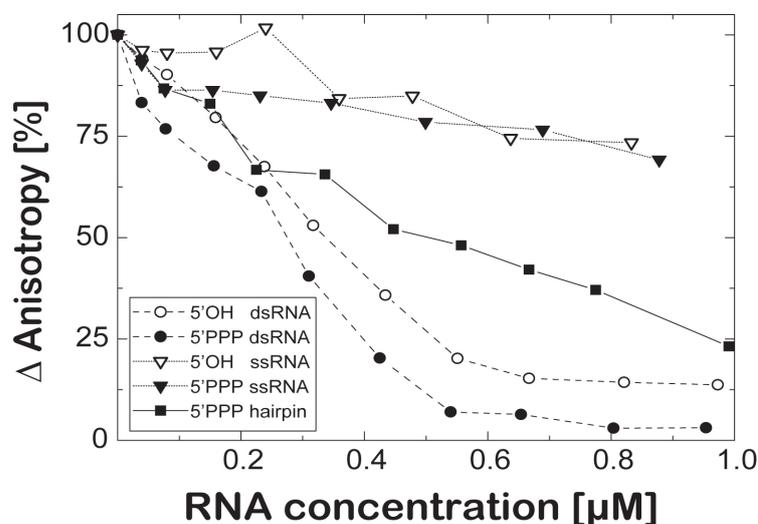


Figure 17 Competition of binding of an AlexaFluor 488-5-U labeled hairpin RNA (*in vitro* transcribed, 40 nM) to LGP2 RD (470 nM) by stepwise addition of different non-fluorescent RNA species (synthetic 5'-OH/5'-OH dsRNA, 5'-PPP/5'-OH dsRNA, 5'-OH ssRNA, 5'-PPP ssRNA and *in vitro* transcribed 5'-PPP hairpin) followed by fluorescence anisotropy. Data points were connected for better outline.

To determine whether 5'-triphosphate is equally important for LGP2 RD - RNA interaction as for RIG-I RD, single and double strand RNA oligonucleotides harboring a 5'-triphosphate were tested for competition. However, no substantial difference in binding of LGP2 RD to the corresponding RNAs with 5'-triphosphate compared to those without was measured.

This contrasts to the 5'-triphosphate dependence of RNA binding by RIG-I RD. Taken together; it appears that RD is an important element in specific dsRNA recognition by LGP2 and that 5'-triphosphates are not central epitopes recognized by LGP2 RD.

3.2.6. The dsRNA Binding Site of LGP2

To specify and characterize the exact RNA binding site of LGP2 RD, previously mentioned point mutations were introduced inside and outside the positively charged cleft that was predicted to be the RNA interacting area (Fig. 16 B).

RNA binding of the wildtype and mutant RD variants was analyzed by fluorescence anisotropy (Fig. 18). Even though it was not entirely clear how many RD binding sites exist on the RNA strands used for anisotropy measurements, the binding isotherm was fitted with a single site-binding model. This was done for reasons of simplification, after trying to fit the data with different binding models and finding that all yield nearly invariant apparent K_d values.

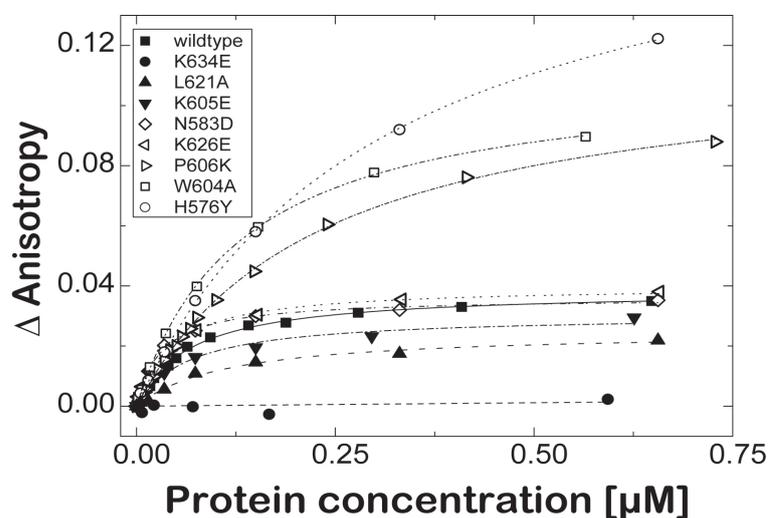


Figure 18 Binding of dsRNA to LGP2 RD. (A) Fluorescence anisotropy changes of a 5'-AlexaFluor 488 labeled 25 bp RNA duplex (37 nM) in response to titration with wildtype (wt) LGP2 RD (filled square, $K_d = 68 \pm 6$ nM) and various mutants, respectively. Control mutation N^{583D} (open diamond, $K_d = 38 \pm 4$ nM), located on the convex site of RD and mutation K^{626E} (open left-facing triangle, $K_d = 51 \pm 5$ nM) do not show significantly altered dsRNA binding affinity. Affinities for mutants L^{621A} (filled up-facing triangle, $K_d = 165 \pm 26$ nM) and K^{605E} (filled down-facing triangle, $K_d = 140 \pm 16$ nM) are slightly decreased, while a mutation of K^{634E} (filled circle) completely suppresses binding. A decrease of binding affinity, but increase in maximum reached anisotropy signal is seen for P^{606K} (open right-facing triangle, $K_d = 230 \pm 10$ nM), W^{604A} (open square, $K_d = 136 \pm 6$ nM) and H^{576Y} (open circle, $K_d = 304 \pm 10$ nM).

3.2.6.1. Study of LGP2 RD's RNA Interaction by Fluorescence Anisotropy

Comparing the obtained results between mutants and wildtype RD it becomes obvious that mutations in the concave, positively charged surface patch (K^{634E}, L^{621A}, K^{605E}, P^{606K}, W^{604A}, H^{576Y}) generally affect RNA binding, while mutations outside this region (N^{583D}, K^{626E}) do not confer any evident alteration in the RNA binding behavior of LGP2 RD. Thus, there is no involvement in RNA binding of the area where SO₄²⁻ ions are coordinated in the LGP2 RD crystal structure.

Point mutation of K^{634E} completely abolished RNA binding of LGP2 RD. This residue is conserved between RIG-I like helicases and situated at the N-terminus of a 3₁₀ turn (η 4) at the center of the concave surface. The equivalent mutation in RIG-I K^{888E} also had a severe effect on RIG-I's RNA interaction and activity (Cui *et al.* 2008).

The conservation of this lysine residue in different RDs with evidently different RNA specificities (5'-triphosphate in RIG-I *versus* dsRNA in LGP2) argues that this site is not involved in distinguishing different RNA epitopes, but more likely is a central core RNA interaction site shared by all three RDs.

The mutant P^{606K} still binds RNA with reasonable efficiency, which might be due to the replacement with a positively charged residue, while the corresponding K^{858A} mutation severely reduced 5'-triphosphate RNA binding of RIG-I RD and abolished RIG-I activity *in vivo*. Other mutations of LGP2 RD in the same region (H^{576Y}, W^{604A}) also bind RNA with reasonable, albeit reduced affinity (Fig. 18). Interestingly, although K_d values for dsRNA binding by these three mutants are slightly lower than for wildtype RD, much higher anisotropy changes are measured compared to the wild type RD. This indicates that the RNA/protein complexes formed by these mutant RD variants are larger than those obtained for wild type RD. This could be due to the fact that these mutant proteins might oligomerize upon binding to RNA. However, such a gain of function by only one single residue change seems unlikely. Another explanation is that these particular mutants can bind simultaneously at multiple sites on the RNA, while the wildtype binds more specifically to only one or a few distinct sites.

3.2.6.2. Electrophoretic Mobility Shift Assays of LGP2 RD – RNA complexes

EMSA with wild type and mutant LGP2 RDs were conducted, to further investigate the divergence of their dsRNA binding affinity and specificity, (Fig. 19). In general, the EMSAs confirm the results of the anisotropy measurements with respect to altered or retained dsRNA binding of specific mutants (Fig. 19 A). Concentration dependent analysis indicates that wild type RD indeed shifts dsRNA to defined, specific bands. An initial shifted species with higher mobility is subsequently converted into a species with lower mobility (Fig. 19 B).

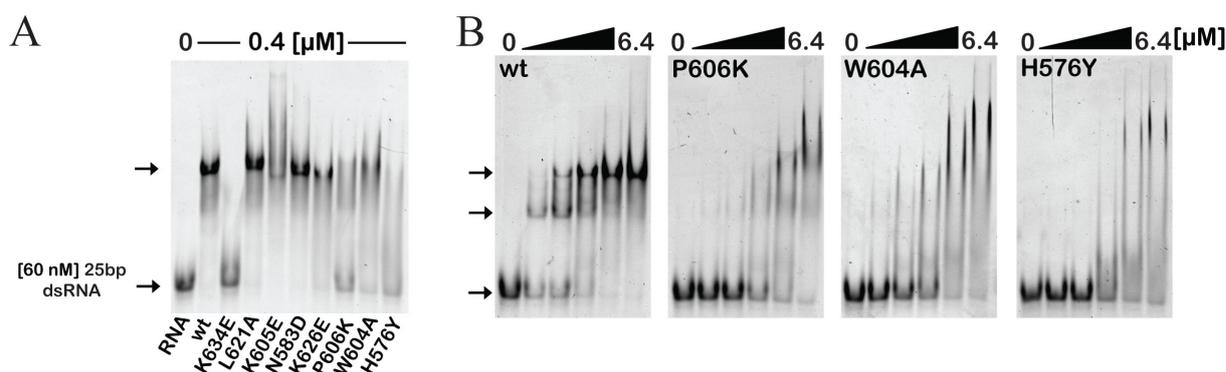


Figure 19 Electrophoretic mobility shift assays of LGP2 RD and mutants with dsRNA. (A) Retardation of 5'-AlexaFluor 488 labeled dsRNA (25 bp, 60 nM) in a 10% native polyacrylamide gel after incubation with 400 nM wt LGP2 RD or indicated RD mutants, respectively. (B) Retardation of 5'-AlexaFluor 488 labeled dsRNA (25 bp, 60 nM) in a 10% native polyacrylamide gel after incubation with increasing concentrations (0, 0.4, 0.8, 1.6, 3.2, 6.4 μ M) of wt LGP2 RD and mutants P^{606K}, W^{604A} and H^{576Y}.

These data suggest that two LGP2 RD molecules can bind to the dsRNA ligand. One possibility is that RD specifically forms dimers on a single site on the dsRNA substrate. Another option is that two binding sites for LPG2 RD exist on the RNA. Since LGP2 RD exhibits higher affinity to dsRNA than to an RNA hairpin of the same concentration and similar stem length, another explanation is that RNA end structures contribute to binding. Hence, the two EMSA species would be corresponding complexes with either one or both RNA ends masked by protein. In any case, the observed species in the EMSAs are well defined, indicating a particularly specific interaction of LGP2 RD with the dsRNA substrate.

Mutant K^{634E} entirely fails to interact with the RNA double strand. Also the distinct bands found for the wildtype RD - RNA complex are lost for the H^{576Y}, W^{604A} and P^{606K} mutants of RD (Fig. 19 B). Instead, an unspecific distribution (“smear”) of slower migrating

complexes can be observed. Aside from the unspecific shifting, the mutants also exhibit lower affinity to the RNA compared to wt RD, indicated by remaining free dsRNA bands for all protein concentrations.

These mutations do not seem to disrupt RNA binding per se, but lower the affinity and possibly lead to a more distributed unspecific binding all along the RNA duplex. These mutant RDs could bind indiscriminately to many possible binding sites on the dsRNA ligand, perhaps also as multimers at higher protein concentrations, which could account for the larger complexes observed in both EMSA and fluorescence anisotropy measurements. Taken together, these data argue that the residues around P⁶⁰⁶ form the specificity site of LGP2 RD.

3.3. RD – RNA Complex Crystallization Attempts

3.3.1. Generation of 5'-Triphosphate RNAs for Co-crystallization with RIG-I RD

In addition to the analysis of LGP2, RNA ligands suited for co-crystallization with RIG-I RD should be established and tested for complex formation and crystallization. At this point a physiological ligand of RIG-I was believed to be 5'-triphosphate ssRNA, despite the helicase domain showing a preference for dsRNA. RIG-I RD had been shown to be the 5'-triphosphate sensor with the binding site being mapped to a positively charged groove on the concave site of the molecule (Cui *et al.* 2008) analog as described for LGP2 RD.

Since no protocols were available for efficient synthesis of 5'-triphosphate containing RNA in quantities large enough for crystallization, *in vitro* transcription had to be employed to generate the 5'-triphosphate ligands. This was achieved by either using recombinant T7 RNA-Polymerase purified in the lab or the Ambion Megashort Transcript kit, which is especially suited for the transcription of the short RNAs that would be required for optimal crystal packing in co-crystallization setups.

Initially, a 58mer of the rabies virus leader (RVL) sequence was produced, according to the one used in former RIG-I RNA binding studies, and tested in analytical gel filtration experiments for binding to either LGP2 or RIG-I RD. Surprisingly, and despite its demonstrated preference for dsRNA, LGP2 RD formed a stable complex with the single-stranded 5'-triphosphate containing 58 nt RVL comparable to RIG-I RD (Fig. 20 A, B). Since the binding was almost as strong as for RIG-I RD, but with LGP2 RD known to exhibit only weak binding towards ssRNA, it could be assumed, that the *in vitro* transcribed sample did not solely contain ssRNA, but possibly also some double strand species or duplex regions. This can result from “back-looping” of the T7 RNA Polymerase when reaching the end of the

template strand rather than producing run-off ssRNA products and RNA-template primed RNA synthesis (Cazenave *et al.* 1994; Arnaud-Barbe *et al.* 1998). Despite the possible inhomogeneity of the RNA, a rather distinct additional, early eluting peak appeared on the elution profile of the gelfiltration for both proteins, suggesting formation of a protein-RNA complex. The number of binding sites on the 58 nt RVL remains unclear for both RDs, with calculated values ranging between 1.3 and 1.7 molecules per RNA molecule, according to the molecular weight standard. Also there is no evidence for possible dimerization of RDs. Furthermore, complex formation with an *in vitro* transcribed ribozyme product with 159 nt was analyzed and showed similar results, whereas incubation of RDs with either synthetic Natriphosphate or a PolyA octamer failed to induce detectable complexes (not shown).

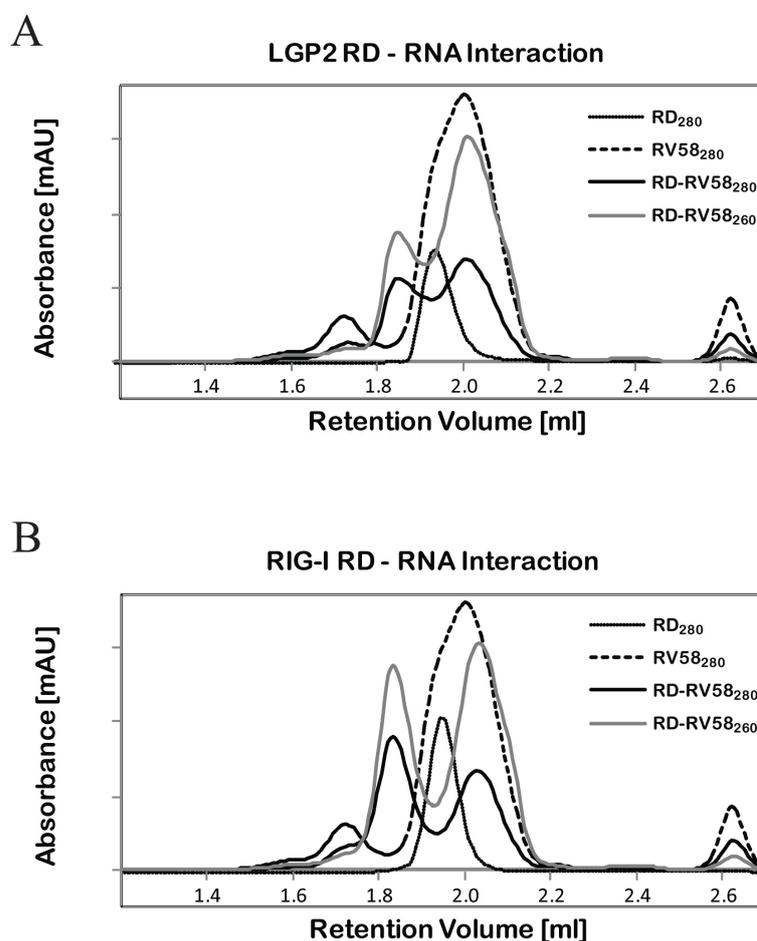


Figure 20 *In vitro* transcribed RNA (58mer RV leader) forms stable complexes with either LGP2 RD (A) or RIG-I RD (B) that are detectable in analytical gelfiltration.

In order to co-crystallize RIG-I RD, shorter RNAs were required, since long uncoated RNA overhangs were assumed to disturb proper crystal packing. Therefore, 5'-triphosphate RNAs with a length around 6 – 21 nucleotides were targeted. Various transcription strategies are shown in Fig. 21.

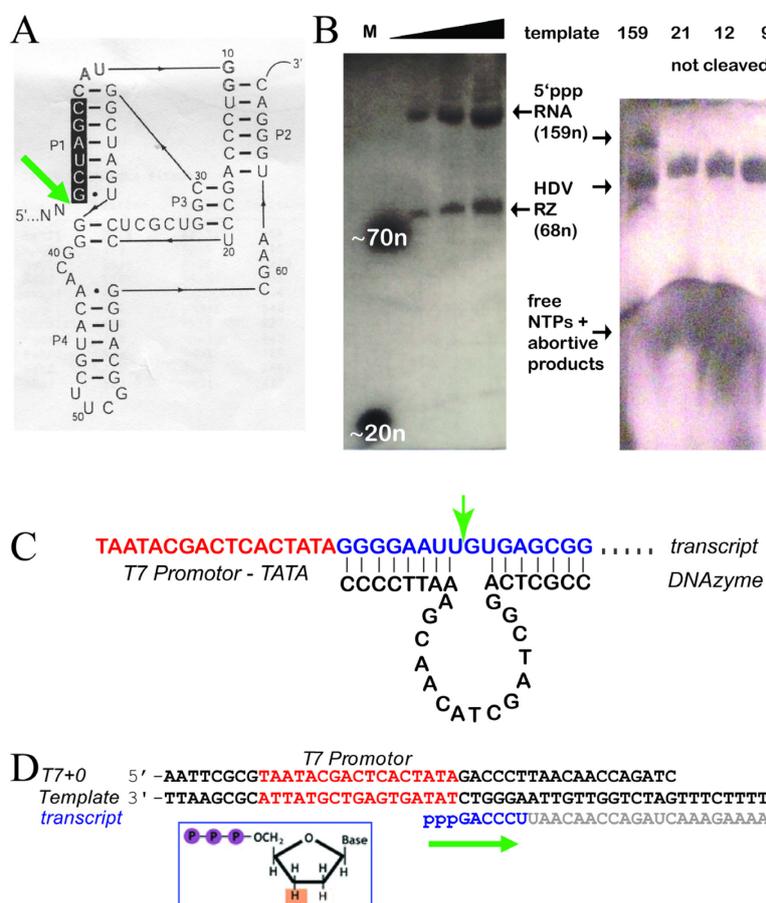


Figure 21 Strategies to obtain short 5'-triphosphate RNAs with homogeneous 3' ends. (A) Secondary structure of a modified HDV ribozyme used. (B) 6M urea/ 12 % (w/v) polyacrylamide gel with increasing yield of 159 nt ribozyme cleavage products after 6h incubation at 37 °C of run-off *in vitro* transcription products obtained with different template concentrations. In comparison ribozyme constructs containing 21, 12 or 9 nt do not appear cleaved. (C) Scheme of a DNAzyme and (D) Transcription termination by incorporation of ddNTPs, such as ddUTP.

Initially ribozyme and DNAzyme mediated generation of *in vitro* transcribed RNA harboring homogenous 3' ends was tested (Fig. 21 C, D). Run-off transcription of long RNA strands is much more efficient than transcription of short targets. This is due to the T7 polymerase's improved initiation and activity on longer template strands. The longer the template, the less premature termination of transcription, usually after the first 2 to 12

nucleotides, occurs. Ribozymes and DNAzymes represent practical tools for obtaining short homogenous RNAs from initially long *in vitro* transcribed strands. In case of the first, the shortness of the sequence to be autocatalytically cleaved off appeared to diminish the efficiency of the reaction. Different from a 159 nt sequence initially used for optimization of reaction conditions, insertions of 9, 12 or 21 nucleotides in front of the ribozyme did not yield homogeneous cleavage products (Fig. 21 B). For the DNAzyme, discrimination between and separation of the DNA and the RNA product and reaction yield proved to be a bottleneck.

As another strategy transcription has been conducted using long templates, but by trying to force synthesis of more abortive short RNA products by using AlexaFluor-labeled UTP, that might potentially promote polymerization stop. This does not necessarily produce a homogeneous product though. A definite polymerization break was achieved by adding dideoxy-UTP instead of normal UTP to the reaction mixture, preventing any further elongation after the first introduced ddUTP. However, the obtained products can still contain a mix of even shorter RNA products if polymerization stops short of the site of ddUTP incorporation.

Even though transcription products were tried to be purified by reversed phase chromatography, samples were usually too dilute and inhomogeneous to yield significant amounts of pure RNA. Furthermore, the inherent proneness of T7 Polymerase and especially the non-commercial one, to early abortion of transcription, leads to a mix of different aborted RNAs between 2 and 12 nucleotides, making targets in the same range hard to purify. Also integration of up to four Gs that are favorable for transcription efficiency at the beginning of the transcript could not significantly improve yields.

Nevertheless, RIG-I RD was able to bind transcription products with incorporated AlexaFluor-UTP or ddUTP, forming a stable complex that could be detected and separated by preparative gelfiltration (Fig. 22). Thus, selection and hence homogeneity of the RNA ligand by RIG-I RD was potentially achieved.

Incubation of RIG-I RD with those RNAs resulted in quantitative complex formation. Comparing the gelfiltration retention volumes of RIG-I RD by itself, residual free RNA oligonucleotide and RIG-I RD in complex with RNA to a molecular weight standard, the complex appears about twice as large compared to RIG-I RD alone. This suggests that two RD molecules specifically interact with the *in vitro* transcribed RNA constructs, possibly *via* the RNA ends as proposed for LGP2 RD. This is speculative, since an exact prediction of the properties of the RNA ligand cannot be made and it is not clear that it really is only the expected short single strand.

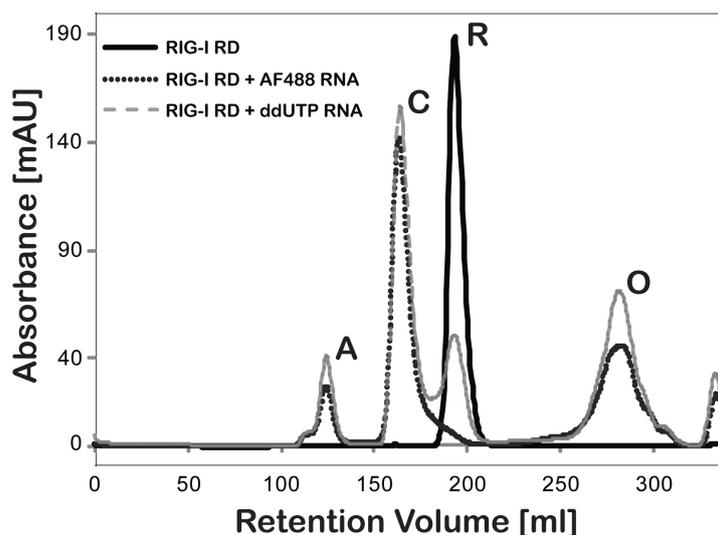


Figure 22 Elution profile of RIG-I RD by itself (solid black line) and in complex with AlexaFluor 488-UTP (dark dotted line) or ddUTP (light dashed line) incorporated RNA from a preparative Superdex S75 26/60 gelfiltration. A: aggregate, O: free RNA oligonucleotide (apparent molecular weight – MW~1.3 kDa), R: RIG-I RD only (MW~15 kDa), C: Complex of RIG-I RD and RNA (MW~36 kDa).

3.3.2. Co-crystallization of RIG-I RD with 5'-Triphosphate RNA

RIG-I RD – RNA complexes were purified by gelfiltration and subsequently screened for crystallization. The best shaped crystals were obtained with the ddUTP-transcript containing samples. Initial screens yielded mainly needles, generally in various conditions containing PEGs and Lithium-salts. Crystals were grown at 4 °C to ensure RNA stability and could be refined to 3-dimensional rods after macro seeding (Fig. 23).

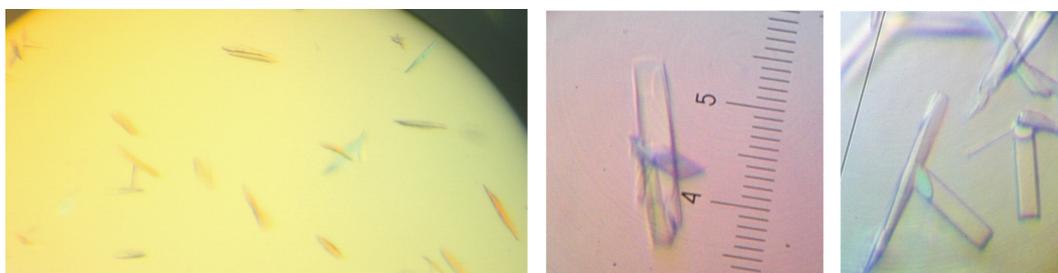


Figure 23 RIG-I RD/RNA complex crystals grown in 100 mM TRIS HCl pH 8.5, 18% (w/v) PEG 8000, 200 mM Li₂SO₄ in Jena Bioscience Classic HTSL Screen I and after refinement.

Unfortunately, all crystals showed weak diffraction (maximum 8 Å) or none at all. It is possible that the RNA quality and homogeneity was not high enough to ensure proper crystal packing and order. The chosen RNA ligand lengths could further not be ideal. Also, the single stranded character of the used RNA ligands might have been disadvantageous, given that later on RIG-I RD became known to actually prefer blunt end double strand RNA containing a 5'-triphosphate as a ligand.

4. AIM2 – Results

4.1. Full Length Mouse AIM2

4.1.1. Identification and Purification of Degradation Products of *mAIM2*

Since human full length AIM2 was prone to aggregation when expressed in *E. coli*, the much better behaving mouse homolog was used for structural and functional studies of AIM2. It will be further on referred to as *mAIM2*. First attempts to purify *mAIM2* with an N-terminal Hexa-His-tag failed due to the protein not binding to Ni-NTA, indicating either inaccessibility of the tag or N-terminal degradation. *mAIM2* could be purified carrying a C-terminal 6xHis-tag or without any tag using Heparin affinity chromatography. The purified protein always exhibited a smaller apparent molecular weight on SDS-PAGE gels than expected and degradation products were detected that could not be resolved by gel filtration chromatography. The best separation of the degraded AIM2 fragments was achieved by cation exchange chromatography (Fig. 24). Without affinity tag, *mAIM2* has a theoretical molecular weight of 40.2 kDa. Purified AIM2 fragments exhibit apparent molecular weights of 34 kDa and 37 kDa according to migration in SDS PAGE. To identify the degradation products of untagged *mAIM2*, samples have been analyzed by mass spectrometry after isolation from distinct SDS PAGE gel bands and tryptic digest. This analysis showed that both major purification products referred to *mAIM2* and that mainly peptide patterns from the N-terminal region were missing in the spectra, also supporting the idea of N-terminal degradation of the protein.

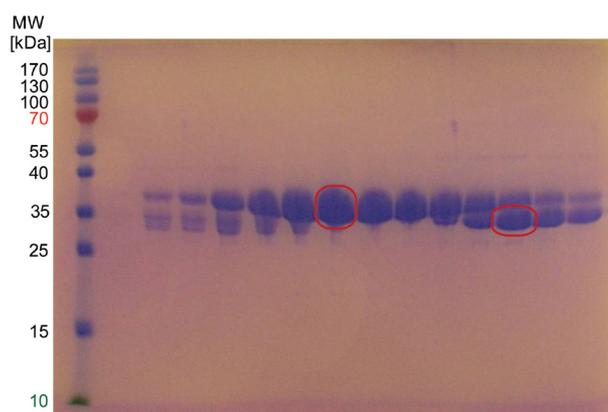


Figure 24 SDS PAGE gel of samples from elution fractions of *mAIM2* from an SP-Sepharose column run as final purification step. Both major bands migrating around 37 and 34 kDa correspond to *mAIM2* according to mass spectrometric analysis (analyzed bands are framed in red).

To finally prove this and determine the exact constructs, Edman sequencing was performed for the two most abundant degradation products. Cleavage sites were identified in either the Pysin domain (residues 1-90, e.g. M75) or between the Pysin and HIN domain in an unstructured region (see chapter 4.1.2, residue T94). Theoretical molecular weights of the corresponding degradation products are even lower than the apparent ones on SDS PAGE.

4.1.2. Structural Model of AIM2

In absence of a stable full length *mAIM2* crystallization target structural modeling has been performed. As verified by various pattern and motive search programs AIM2 was found to contain an N-terminal death domain fold Pysin (or PAAD/DAPIN) domain (PYD, res 1-90) and a C-terminal HIN-200 (or IF120x) domain (res 140 – 354). Sequence alignments were performed separately for the domains (Fig. 25). Through the PFAM database, homolog domains with known 3D structures were identified and best matches for each domain were then used to model the tertiary structure of *mAIM2* with Modeller (Fig. 26 A). For the DNA binding HIN domain the two HIN domains of human Gamma-Interferon-Inducible Protein 16 (Ifi16, PDB: 2OQ0, 3B6Y) exhibit high homology with sequence identities over 40% (Fig. 25 B). Both structures were used for AIM2 modeling. Pysin domains of AIM2 are not as highly homologous to other Pysin domains such as PYD of *hASC*, human and mouse NACHT-, LRR- and PYD-containing protein (*hNALP1* and *mNALP10*, part of NALP inflammasomes), human Myeloid cell nuclear differentiation antigen (*hMNDA*) and mouse Interferon-Inducible Protein 205 (*mIfi205*), for which structures are available (Fig. 25 A). Modeling was performed based on the *mNALP10* (PDB: 2DO9) PYD structure that shows highest similarity, amongst the candidates, to the *mAIM2* PYD with a sequence identity of ~25%.

The region between the Pysin and HIN domain of *mAIM2* could not be reliably modeled due to a lack of homology models and as it does not belong to a distinct domain. Secondary structure predictions indicate that this region is unstructured (Fig. 25 C). Sheets and helices predicted by various algorithms (like JPred and PSIPred) in this area have a much lower probability or prediction confidence than nearby structural elements.

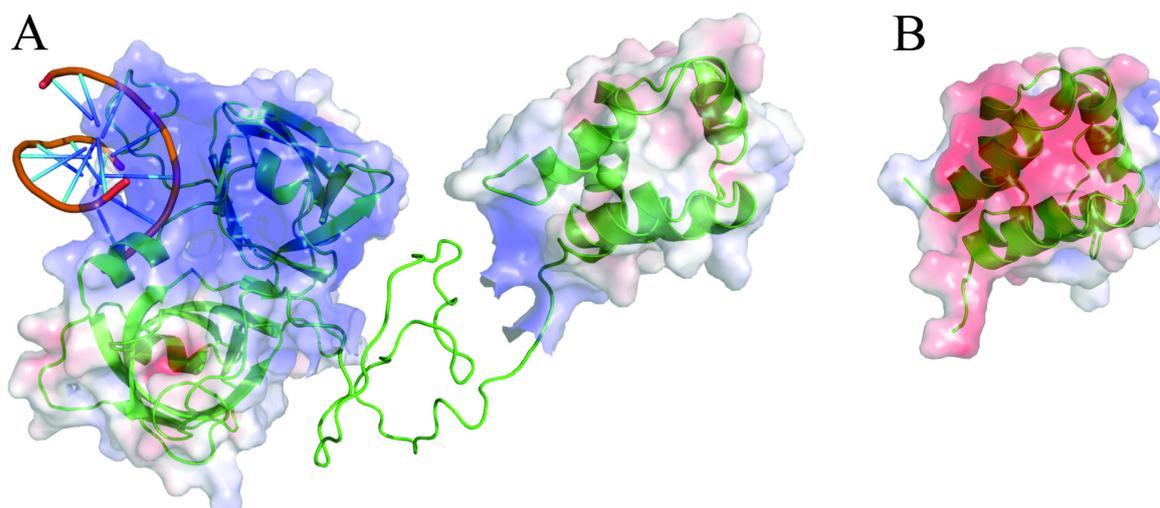


Figure 26 Structural model of (A) *mAIM2* with the N-terminal Pyrin domain on the right and the C-terminal HIN domain depicted on the left, connected by a flexible linker region. The tentative electrostatic potential ranges from -5 kT (red) to 5 kT (blue) and a dsDNA has been docked to a positively charged groove between the two conserved OB-folds of the HIN domain. In comparison (B) shows the experimentally solved structure of the ASC-Pyrin domain (PDB: 2KN6) harboring the same death domain fold as the modeled AIM2-PYD.

4.2. The AIM2 HIN Domain

Since the Pyrin domain of AIM2 is prone to degradation, and therefore difficult to work with, the DNA-binding HIN domain by itself was focused on for structural and functional characterization.

HIN constructs were chosen according to sequence alignments, fold predictions, mass spectrometric results from degradation products of full length AIM2 preparations that suggest an N-terminal degradation and based on Edman sequencing results also from degradation products. Constructs used are *hAIM2* 140-343, *mAIM2* 146-354, *mAIM2* 94-354 and *mAIM2* 137-354.

The HIN domain of human AIM2 containing residues 140 – 343 (native C-terminus) could be expressed in soluble form with an N-terminal Hexa-His-tag in *E. coli* Rosetta (DE3) cells, yielding ~35 mg protein from a 4.5 l expression. Removal of endogenous DNA bound to the HIN-domain was achieved by a 2M NaCl high salt wash of the protein while immobilized on Ni-NTA. Crystallization screens of the purified construct without DNA did not result in any hits and addition of stoichiometric amounts of DNA resulted in immediate precipitation of the protein. EMSAs with quite small amounts of DNA could show binding of the HIN domain to dsDNA, however in a weak, unspecific manner, possibly caused by the constructs proneness to aggregation in the presence of DNA (not shown).

The HIN domain of mouse AIM2 (residues 146-354) could be purified similarly and appeared more stable. After removal of the His-tag the construct was less prone to aggregation when incubated with DNA. Since HIN constructs generally seemed to be more stable after removal of the His-tag, new constructs (AIM2-HIN: 146-354, 94-354, 137-354) were cloned and purified without any tag. Sufficient purity of the DNA-binding constructs was thereby achieved by purification using Heparin affinity, cation exchange and size exclusion chromatography.

4.3. Evaluation of mAIM2 – DNA Complex Formation by Electrophoretic Mobility Shift Assays

To evaluate *mAIM2* binding to short DNAs and the contribution of the HIN domain EMSAs have been carried out using “full length” *mAIM2* (protein corresponding to the sample with lowest mobility in SDS PAGE from the preparation of the full length construct) and the HIN domain construct *mAIM2* 146-354 (Fig. 27). Two different DNA lengths and topologies were tested for binding, an 11 bp + 4 nt hairpin and a 35 bp DNA double strand.

The 11 bp hairpin can accommodate binding of two HIN domain molecules as can be seen by the successive appearance of two distinct bands with lower mobility than DNA itself in the EMSAs. For the longer AIM2 construct there seems to be only one binding site on this short DNA, judging from one lower mobility band appearing with higher protein concentrations. This is likely due to steric or spatial constriction. With increasing AIM2 concentration the entire sample does not migrate into the gel, which might be due to formation of higher oligomeric assemblies or just unspecific aggregates that are above the molecular cutoff of the gel matrix (Fig. 27 A). This was also observed for the longer DNA construct. Furthermore it seems that the 35 bp DNA can facilitate binding of two molecules of either the HIN domain or also the “full length” construct, resulting in two distinctly shifted bands (Fig. 27 B). Binding affinities can only be estimated from the EMSAs but seem rather high. The apparent K_d is around 100 nM and even though the number of molecules bound to the different DNAs varies, affinities appear to be similar.

It should be noted that although the fast annealing procedure (5 min at 95°C then immediate storage on ice) used for the hairpin DNA should ensure a proper hairpin formation, a self-annealing of the 26 nt DNA that would result in a 26 bp dsDNA rather than the 11 bp hairpin, cannot be entirely ruled out. Hence, no reliable conclusion concerning the minimum dsDNA length required for AIM2 or only HIN domain binding can be derived.

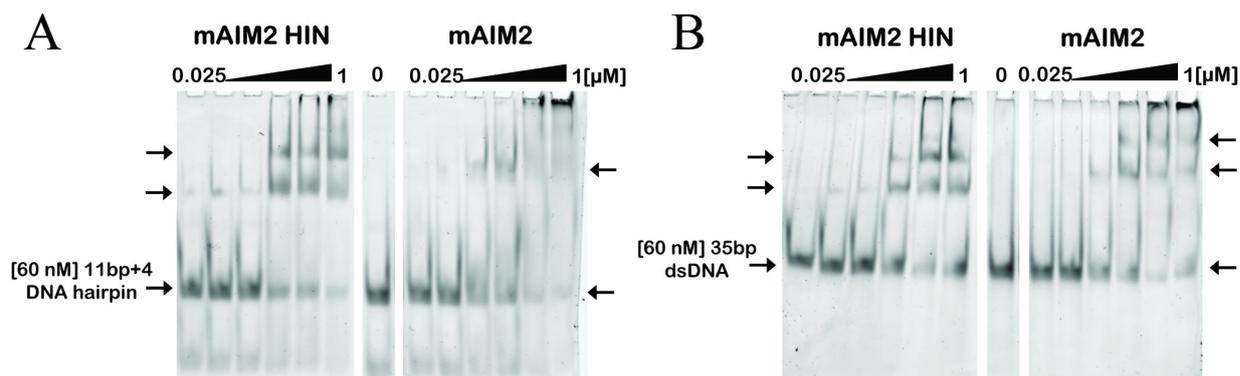


Figure 27 Electrophoretic mobility shift assays of *mAIM2* full length (theoretically) and HIN (146-354) domain constructs studying binding to (A) a DNA hairpin with 11 bp stem and a tetraloop labeled with 3'-FAM or (B) a 35bp annealed dsDNA containing 5'-ATTO488 as fluorescent label. Protein concentration steps used are: 0, 0.025, 0.05, 0.1, 0.25, 0.5 and 1 μ M with a constant DNA concentration of 60 nM.

4.4. Analytical Gelfiltration of Complexes of *mAIM2* and DNA Ligands Suited for Crystallization

Prior to setting up crystallization screens of AIM2-DNA complexes, potential ligands were analyzed for quantitative binding to *mAIM2* by analytical gelfiltration (Fig. 28). Amongst the four DNA samples tested an 8 bp dsDNA was the only one failing to form a stable, detectable complex with the protein ("full length" *mAIM2*). This might either be due to AIM2 requiring more than 8 bp in order to bind efficiently or more likely to instability of this short DNA fragment at room temperature. On the other hand three DNA hairpins, with varying stem length (11, 18 and 25 hp; Table 7) and harboring a tetraloop, that were derived from the successfully crystallized domain E of *Thermus flavors* 5S rRNA (PDB: 361D9) appear to be quantitatively bound. Formation of the protein DNA complex was very efficient, resulting in an obvious additional, early eluting peak in gelfiltration, while the initial peak corresponding to the DNA by itself was entirely lost.

Stoichiometries calculated from these elution profiles suggest a single binding site for *mAIM2* on either the 11 or 18bp hairpin, while more molecules (calculated: 3.5) appear to bind to the 25 bp hairpin (Fig. 28). This is in agreement with previously shown EMSAs.

Interestingly, *mAIM2* exhibits a higher molecular weight in analytical gelfiltration than in SDS-PAGE.

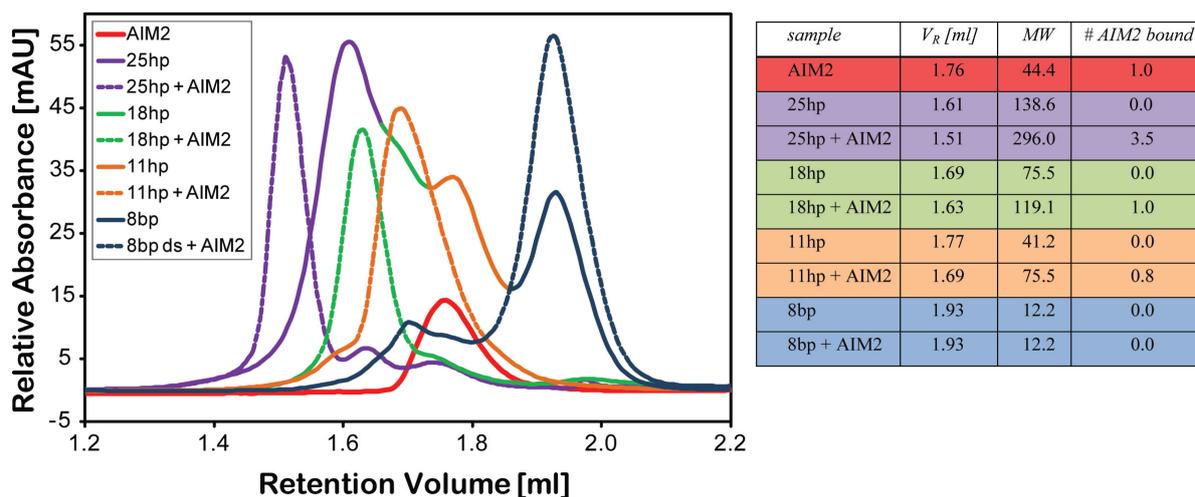


Figure 28 Chromatograms of analytical gel filtration runs of *mAIM2* alone and in complex with different RNAs (8 bp dsRNA, 11+4 hp, 18+4 hp, and 25+4 hp) and calculations for binding stoichiometries derived from them. Molecular weights were calculated according to a run of a molecular weight standard (theoretical MWs: *mAIM2* ~ 40.2 kDa, 11 hp ~ 8 kDa, 18 hp ~ 12.3 kDa, 25hp ~ 16.6 kDa, 8 bp ~ 4.8kDa).

4.5. Crystallization of AIM2 – dsDNA Complexes

For crystallization and binding assays the longest stable *mAIM2* fragment (~37 kDa), obtained during purification without any tag was used, initially. For co-crystallization DNA hairpins of different lengths (11 bp+4 nt, 18 bp+4 nt and 25bp+4 nt) that had been shown to bind to AIM2 and an 8 bp dsDNA, that failed to form a detectable complex with AIM2 in gel filtration, were utilized (Table 7). Protein was used with a concentration of ~10 mg/ml and DNA was initially added in stoichiometric amounts.

Crystals containing the long *mAIM2* construct were predominantly obtained in presence of the 11 or 18 hp (some for 25 hp and none for the 8 bp dsDNA), in the PEG-based Jena Classics HTS I screen as well as the Qiagen Nucleix Suite (Table 10). Setups with solely protein or DNA in the same conditions resulted in clear drops or precipitate, respectively, suggesting that the crystals contained a complex of both.

Table 10 Examples for crystallization hits in initial screens for the complex of either full length *mAIM2* or HIN domain constructs containing residues 146-354 with an 11 or 18 bp hairpin DNA.

<i>mAIM2</i>							
Jena Cl I	Precipitant I [%] w/v	Precipitant II [mM]	Precipitant III [mM]	Buffer [mM]	pH		
C 11	PEG 4000 30	CaCl ₂ 200		HEPES Na Salt 100	7.5		
C 12	PEG 4000 30	Na Acetate 200		TRIS-HCl 100	8.5		
E 9	PEG 4000 32		LiCl 800	TRIS-HCl 100	8.5		
F 7	PEG 6000 28		LiCl 500	TRIS-HCl 100	8.5		
F 8	PEG 6000 30	Na Acetate 100	LiCl 1000				
Nucleix	Precipitant I [%] w/v	Precipitant II [mM]	Precipitant III [mM]	Buffer [mM]	pH		
H12		MgCl ₂ 10	Spermine 50	Na Cacodylate 50	6.5		
<i>mAIM2</i> HIN domain (146-354)							
Jena Cl I	Precipitant I [%] w/v	Precipitant II [mM]	Precipitant III [mM]	Buffer [mM]	pH		
C 7	PEG 4000 25	MgCl ₂ 200		MES Na Salt 100	6.5		
C 8	PEG 4000 25	CaCl ₂ 200		TRIS-HCl 100	8.5		
C 11	PEG 4000 30	CaCl ₂ 200		HEPES Na Salt 100	7.5		
E 9	PEG 4000 32	LiCl 800		TRIS-HCl 100	8.5		
Jena Cl II	Precipitant I [%] w/v	Precipitant II [%] w/v	Precipitant III [mM]	Buffer [mM]	pH		
D 4	MPD 50	NaCl 50	iso-Propanol 20	Na Acetate 50			
D 7	MPD 60	CaCl ₂ 10		Na Acetate 100	4.6		
D 8	MPD 70			MES Na Salt 100	6.5		
D 9	MPD 70			TRIS- HCl 100	8.5		

Unfortunately, refined crystals for full length *mAIM2* – DNA complexes from hanging or sitting drop grew either to the plate or in the PEG skin of the drop. These crystals were very fragile and were not able to be loop mounted successfully (Fig. 29).

Edman sequencing results for degradation products of the purified full length *mAIM2* had suggested that the protein construct used for crystallization might contain the C-terminal HIN domain and a part of the unstructured region connecting it to the Pyrin domain. Therefore, for further crystallization attempts only the HIN domain constructs comprising residues 146-354, 94-354 and 137-354 that were not as prone to further degradation were used.

For co-crystallization with the HIN domain constructs only the 11 and 18 bp hairpins were used and different protein to DNA ratios were tested (1:1, 1:2, 1:3 and 1:4). Also co-purification of the protein-DNA complex by gel filtration has been applied prior to crystallization setups. Complexes containing an 11 bp hairpin as DNA ligand yielded the most hits in initial crystallization screens.

While again many of the conditions favoring crystal growth contained high PEG4000 conditions, some additional MPD conditions were found in the Jena Classics HTS II screen as

well as the Qiagen MPD suite (Table 10). However, even though a slight improvement in crystal growth was observed when the reservoir size was increased, crystals did not grow to reasonable sizes in refinement setups. In addition, heavy precipitation occurred immediately after the drop setup, making seeding methods difficult.

None of the various hit conditions from crystallization screens was able to be refined to produce crystals of sufficient quality to be measured. Additionally, setups at different temperatures, varying protein concentrations, increasing reservoir volumes, supply of additives, addition of TCEP, use of Selenomethionine protein (full length construct), seeding and changes in constructs were tested but did not result in improved crystals.

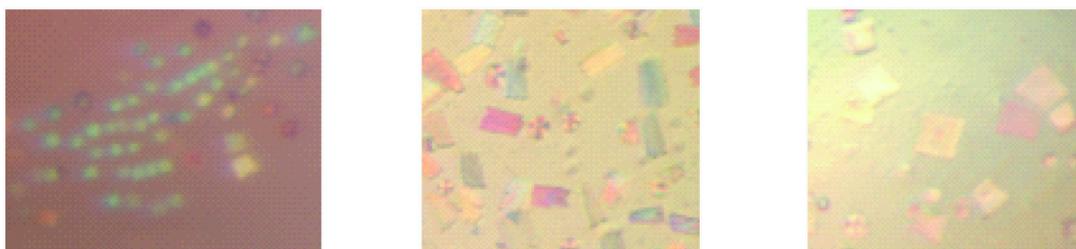


Figure 29 Crystals of *mAIM2* in complex with an 11 bp+4 nt DNA hairpin in refinements from Jena Classics I screen, condition F7 or E9, containing 30 % PEG 6000/0.5 M LiCl/0.1 M TRIS pH 8.5 or 30 % PEG 4000/0.8 M LiCl/0.1 M TRIS pH 8.5, respectively.

4.6. *AIM2* and *ASC* Interaction

ASC and its Pyrin domain have been reported to be prone to aggregation when purified due to their bipolar character. This was also observed with various ASC-constructs and tags. An N-terminal GST-fusion construct of mouse ASC could be purified successfully. Since full length ASC protein contains a flexible region between its PYD and CARD, the main focus was on purifying PYD-only constructs of *mASC* and *hsASC*, as they were anticipated to be best suited for structural studies of a ternary DNA-*AIM2*-ASC(PYD) complex.

The GST-tagged PYD constructs were equally soluble and removal of the GST-tag by PreScission protease digest of the mouse construct in a small scale yielded protein in the soluble fraction (Fig. 30), independent of the presence of *AIM2*. Unfortunately, in larger scales, PYD was aggregating after PreScission treatment in either dialysis, batch or on a GSH-Sepharose column.

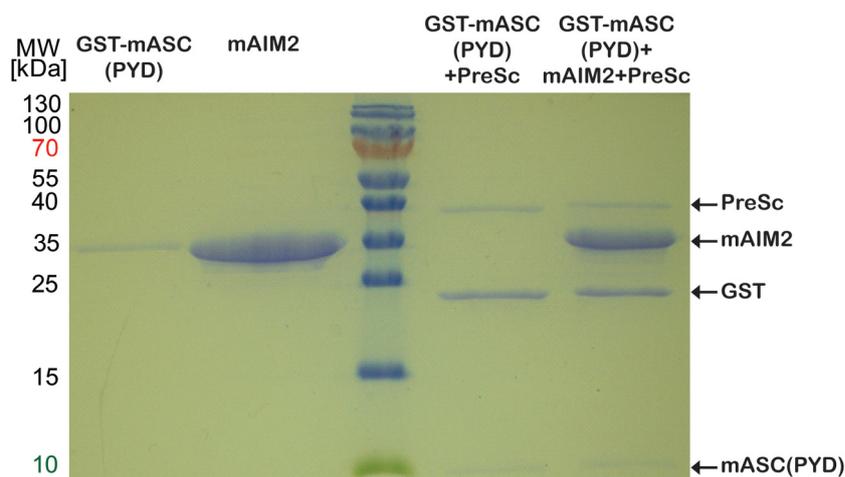


Figure 30 Soluble fractions of PreScission treated samples of GST-*mASC*(PYD) in small scale in presence and absence of AIM2.

Due to the difficulties in purifying ASC's PYD and the degradation of AIM2 in its PYD region, that is required for homotypic PYD interactions, testing of complex formation of the two proteins and the DNA dependence of this event proved challenging.

Pulldown experiments were carried out with GST-tagged *h/mASC* full length or PYD constructs and *h/mAIM2* constructs containing a His-tag (Fig. 31). The two proteins were either co-expressed (from different, but compatible vectors), co-purified (mixing of pellets prior to cell lysis) or purified separately. In the latter case, AIM2-6xHis was first purified on a Ni-NTA column and the elution fraction was added to GSH-Sepharose resin already pre-loaded with GST-ASC.

Pulldown samples were first analyzed by SDS PAGE, which however only gave a vague indication of the interaction of recombinant AIM2 with ASC or ASC-PYD. While the GST-tagged ASC constructs were easily distinguishable (compare Fig. 31, Coomassie stained gel lane G and P), AIM2 appeared degraded (Fig. 31 Coomassie stained gel lane H). Furthermore, an impurity with an approximate apparent molecular weight in the range of full length AIM2 made its detection ambiguous.

Tag-specific immunoblotting of the samples verified a binding of *mAIM2* to *mASC*-PYD(1-93) (Fig. 31). The interaction was detected for the pulldown of purified *mAIM2*-6xHis by immobilized GST-*mASC*-PYD(1-93), as well as in case of co-purification and co-expression of the two constructs by GSH-affinity chromatography. Human constructs and full length ASC samples were not immunoblotted but judging from SDS-PAGE analysis (not shown) an interaction appears equally likely.

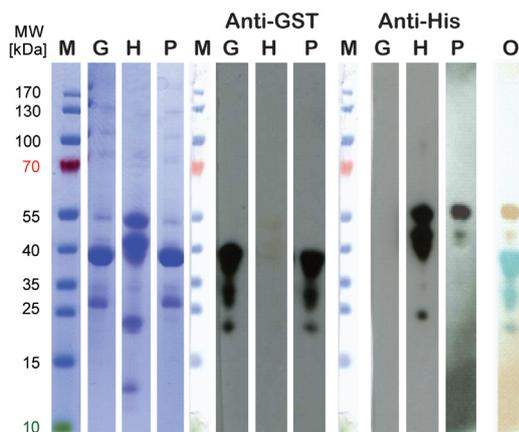


Figure 31 Pull-down interaction assay of *mAIM2* (C-terminal 6xHis-tag) and *mASC-PYD* (1-93; N-terminal GST-tag). GST-*mASC-PYD* was immobilized on GSH-Sepharose. Samples of protein loaded GSH- or Ni-NTA beads in SDS PAGE or Western blot detected by either a GST- or His-tag specific antibody are shown. (M: molecular weight standard; G: GST-*mASC-PYD*; H: *mAIM2*-6xHis; P: pull-down, Ni-NTA eluted *mAIM2*-6xHis bound to immobilized GST-*mASC-PYD*; O: overlaid pull-down lanes, red - anti-His, blue - anti-GST)

The smaller, faster migrating *mAIM2* degradation products are lost in the pull-down sample (compare Fig. 31 anti-His H and P) compared to the Ni-NTA elution sample of *mAIM2*. Considering that the detected His-tag is C-terminal and the earlier shown N-terminal degradation of AIM2 this proves that the N-terminal AIM2-PYD is required to facilitate ASC-PYD interaction. However, a significant amount of *mAIM2* species lagging parts of the N-terminal PYD can still be pulled down by ASC-PYD, which implies the presence of a minimum required binding site in the more C-terminal part of AIM2-PYD.

A high absorbance at 260 nm in an UV/visible spectrum of *mAIM2*, used in the pull-down assay, indicates that cellular DNA remains bound to its HIN domain during the purification steps. DNA might aid AIM2/ASC complex formation and its dependence needs to be further assessed. This assay also shows that a fraction of undegraded AIM2 can be obtained if only an affinity purification step is applied. The presence of ASC-PYD might even stabilize AIM2 and prevent its entire degradation. The apparent molecular weight of the presumed full length *mAIM2* appears higher than the theoretical one.

The verification of the complex formation between AIM2 and ASC is a prerequisite for the identification of minimal interacting constructs as crystallizable inflammasome subcomplexes. Moreover, this finding makes fusion proteins of AIM2 and ASC or ASC-PYD valid crystallization targets.

To further improve the stability of AIM2/ASC complexes, future expression of the single components or fusion proteins in insect cells might be of advantage.

A list of cloned constructs and fusion proteins is shown in Table 11. Most of the already tested fusion-constructs could be expressed and were initially soluble; but unfortunately they were lost during purification due to aggregation. Optimization of purification conditions could help to improve this.

More of the already cloned fusion-constructs (Table 11) need to be tested for soluble expression and stability during purification. Expression in a baculovirus / insect cell system was not attempted for the single or fusion constructs. AIM2 expressed in insect cells however, could be more stable and the introduction of post-translational modifications or phosphorylations, even though not mammalian like, could improve complex formation between AIM2 and ASC.

Table 11 List of hitherto cloned constructs of AIM2 and ASC from mouse and human and fusion variants with respective vectors. A “+” indicates the existence of the construct, with green background indicating good solubility, yellow for limited solubility and red for insoluble expression right away. White “+” fields depict existing clones that yet need to be tested for expression and solubility.

		pFBDM (6His- PreSc)	pET21a (STOP)	pET21a (C-6His)	pET28 (N-6His)	pET28 (N-6His+ SUMO1)	pGEX6P2 (N-GST)	pGEX4T-1 (N-GST)
AIM2	mm	+	+	+	+	-	-	-
	hs	+	+	+	+	+	+	-
AIM2 (HIN-200)	mm(94-354)	-	+	-	-	-	-	-
	mm(137-354)	-	+	-	-	-	-	-
	mm(146-354)	-	+	-	+	-	-	-
	hs(140-343)	-	+	-	+	-	-	-
ASC	mm	-	-	-	-	+	+	-
	hs	-	-	-	+	-	-	-
ASC (PYRIN)	mm(1-93)	-	+	-	+	+	+	-
	hs(1-92)	-	+	-	+	-	+	-
ASC(PYRIN)_20 (PreSc)_AIM2	mm	+	+	-	+	-	-	-
	hs	-	-	-	-	-	+	-
ASC(PYRIN)_11 (PreSc)_AIM2	mm	+	-	-	+	+	+	+
	hs	-	-	-	-	-	+	-
ASC_20(PreSc)_ AIM2	mm	-	-	-	+	-	-	+
	hs	-	-	-	-	-	-	+
ASC_11(PreSc)_ AIM2	mm	-	-	-	-	-	+	+
	hs	-	-	-	-	-	-	-

5. Discussion

Pattern recognition receptors of the innate immune system that are responsive to nucleic acids must be especially sensitive and specific when it comes to recognition of their particular pattern. They have to be able to strictly discriminate between intrinsic and pathogenic nucleic acids to prevent false signaling and auto-immune reactions. Their cytosolic localization represents a prerequisite for concise pattern detection. Furthermore, there is a need for a range of different receptors that are highly specialized for sensing certain subtle modifications or patterns characteristic of pathogenic nucleic acids only.

RIG-I like receptors represent a group of cytosolic helicases that facilitate the detection of viral or virus derived RNA motives and trigger pro-inflammatory signaling in response. For example, RIG-I has been shown to discriminate RNAs by the presence of a 5'-triphosphate, a common pattern in replicating viruses. Since intrinsic RNAs are usually processed or capped this pattern is solely pathogen associated.

It could be confirmed that the C-terminal domain is indeed regulatory rather than repressing and confers pattern recognition in RLRs. This is corroborated by locating and characterizing the RNA interaction sites in RDs and by emphasizing differences in molecular details that contribute to differing pattern specificity amongst the group. In particular, the crystal structure of the RD of LGP2 was solved and extensive analysis of its RNA-binding behavior was conducted. Thus, a better understanding of the regulatory mechanism that LGP2 exhibits on RLR signaling was gained.

With AIM2, an inflammasome forming PRR directed against DNA was analyzed. In addition, its interaction with DNA and the inflammasomal adaptor ASC was studied. Constructs and DNA ligands could be identified that appear suitable for crystallographic analysis. Thus, strong progress towards the determination of molecular structures of AIM2 inflammasome subcomplexes has been made.

5.1. RLR Regulatory Domains Have a Common RNA Binding Site

A comparison of the solved crystal structure of LGP2 RD with structures of the RDs of RIG-I (PDB: 2QFD) and MDA5 (PDB: 3GA3) reveals the extremely high structural conservation of this domain amongst the group of RLRs (Fig. 32 A).

The crystal structure of LGP2 RD combined with functional analysis exhibits a highly positively charged cleft on the concave surface that confers RNA ligand binding, similar to

RIG-I RD. A single conserved Lysine residue in the center of this positively charged patch has been proven to be indispensable for RNA interaction of LGP2 RD. This K⁶³⁴ corresponds to K⁸⁸⁸ in RIG-I that has been established as equally crucial.

The MDA5 RD structure was initially modeled (Pippig *et al.* 2009) but more recently an experimental structure (PDB: 3GA3) (Li *et al.* 2009a) was released. Both MDA5 RD structure and model reveal a positively charged concave surface similar to the RD of RIG-I and LGP2. Thus, the area can be highlighted as an important RNA recognition site in RLR RDs. In addition, the conserved lysine residue, K^{984MDA5}, found to be essential for RNA binding by RIG-I and LGP2 RD, is situated at the same position as K^{888RIG-I} and K^{634LGP2} (Fig. 32 B).

The results allow to postulate a common core RNA binding site for the region around this conserved lysine and 3¹⁰ helix η 4 in the positively charged groove of RLR RDs.

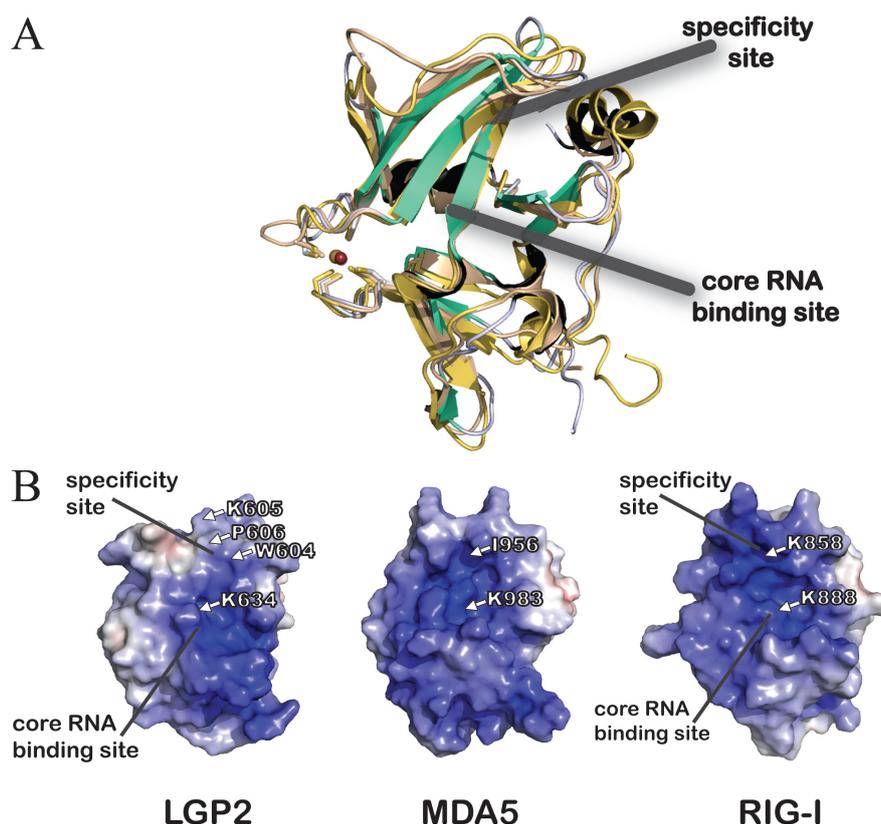


Figure 32 (A) Superposition of the crystal structures of RLR RDs with the proposed RNA ligand specificity and core RNA binding site highlighted. (LGP2 RD: green sheets, black helices, grey loops; RIG-I RD: pale orange; MDA5 RD: pale yellow). (B) Comparison of the surface charge potential of the concave side of RLR RDs with important residues and the RNA interaction sites highlighted (PDB: 2W4R, 3GA3, 2QFD).

Recent structures of RIG-I (PDB: 3LRR, 3NCU) and LGP2 RD (3EQT) in complex with blunt end dsRNA and a 5'-triphosphate in the case of RIG-I correlate with this finding. The conserved lysine forms hydrogen bond contacts with the non-bridging phosphate oxygens of the α phosphate in RIG-I RD, while the corresponding lysine in LGP2 appears to be in hydrogen bonding distance for either the analog terminal backbone phosphate or the subsequent one (see chapter 5.1.2., Fig. 33 B, C).

5.1.1. RLR RNA Binding Specificities are Determined by a Variable Loop Region in the RD

Despite the common RNA binding site, LGP2 has been shown to exhibit different ligand specificity to RIG-I with the regulatory domain conferring discrimination between certain RNA species. While the RIG-I RD ligand was initially thought to be nonspecific 5'-triphosphate containing RNA, it has recently been more accurately determined to be blunt end dsRNA containing a 5'-triphosphate (Schlee *et al.* 2009; Schmidt *et al.* 2009). In contrast, LGP2 RD also preferably binds blunt end dsRNA, however this is entirely independent of the presence of a 5'-triphosphate moiety.

Mutational studies of LGP2 RD suggest that an area centered on β -sheet 6 that is directly adjacent to a flexible loop and harbors various unconserved but functionally important residues, including W⁶⁰⁴ and P⁶⁰⁶ confers selectivity in binding different RNAs. Also MDA5 differs from RIG-I and LGP2 in this region of the RD. It for example possesses an isoleucine (I^{956MDA5}) in the place of P^{606LGP2} or K^{858RIG-I}, thereby the presence of the proposed ligand specificity site is underpinned. Nevertheless, for MDA5, the nature of the PAMP likely sensed by its RD remains to be further investigated. Its ligand specificity might correlate more with LGP2, given the absence of an additional positively charged triphosphate binding pocket (K⁸⁴⁹/K⁸⁵¹) that is only present in RIG-I but not the other RDs.

5.1.2. LGP2 RD Binds to dsRNA Ends

From EMSAs and fluorescence anisotropy titration experiments using wildtype LGP2 RD and point mutated variants a specific binding of LGP2 RD to blunt double strand RNA ends can be concluded. This is based on the appearance of two differently sized distinct RNA shifting complexes in EMSAs, while a dimer formation of LGP2 RD in presence of RNA was never observed. Also, in fluorescence anisotropy competition assays an RNA double strand with two blunt ends was bound preferentially over a hairpin RNA of similar stem length. Furthermore, LGP2 RD's binding to RNA exhibits no obvious sequence specificity, making

the presence of two such specific binding sites along the RNA double strand very unlikely. If more RD molecules are able to non-specifically bind along the RNA rather than at the ends, a less distinct complex formation would be expected, as was found for RDs with point mutations in the proposed specificity site.

This proposal is supported by a recently solved structure of LGP2 RD in complex with a hexameric dsRNA (3EQT) (Li *et al.* 2009b). Superposition of the apo-protein with the RNA bound version does not show any significant conformational changes. Both molecules are very similar as can be seen by the virtually identical Zn²⁺-coordination site (Fig. 33 A). Closer examination of the RNA-interacting residues, that we identified by mutational studies, only exhibits minor changes. While H⁵⁷⁶ has the entirely same position in the superposed molecules and K⁶³⁴ only slightly differs in its sidechain orientation, W⁶⁰⁴ is more exposed in the complex structure. This allows for RNA interaction as opposed to a buried orientation in apo-LGP2 RD. All three residues are within hydrogen bonding distance to either the phosphate backbone (K⁶³⁴) or sugar oxygens (H⁵⁷⁶, W⁶⁰⁴) of the RNA ligand (Fig. 33 B).

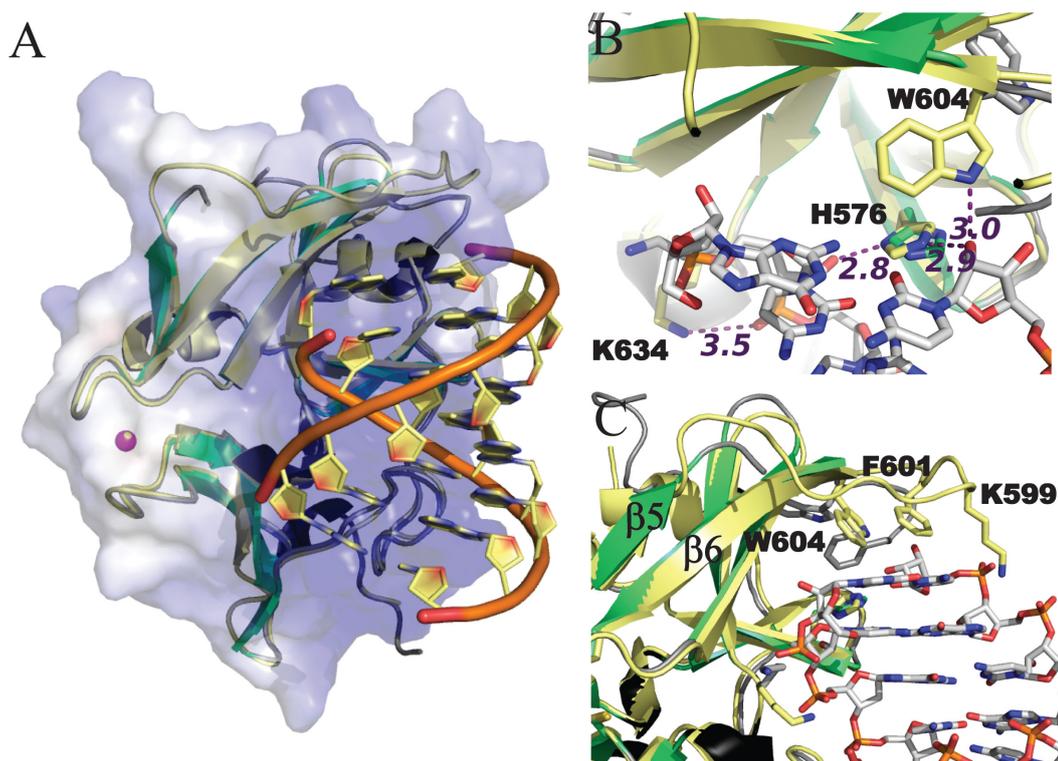


Figure 33 (A) Superposition of apo LGP2 RD (2W4R; green, black, grey) with the RNA bound molecule (3EQT; pale yellow). Electrostatic potential shown for apo LGP2 RD ranges from -5 kT (red) 5 kT (blue). (B) Comparison of single residues of LGP2 RD in RNA bound and apo state with likely hydrogen bonds between protein side chains and RNA. (C) Possible lid-like interaction of the unstructured loop region between $\beta 5$ and $\beta 6$ with bound RNA.

Another conclusion from the initial LGP2 RD structure was that the region between β -sheets 5 and 6 forms a loop that is not defined in the electron density due to its flexibility. This loop appears to shield the specificity side around W⁶⁰⁴ from the surface by slightly burying it. Additionally, in the complex structure the loop appears to form a lid on top of the dsRNA. Flexibility in this region therefore makes sense, to allow RNA to enter the binding site and then be retained by “closing the lid”. Further residues in this region such as F⁶⁰¹ or K⁵⁹⁹ seem capable of interaction with the RNA by either hydrogen bonds or base stacking (Fig. 33 C). Interestingly, the flexible loop region between β -sheets 5 and 6, located in close proximity to the ligand specificity site around P^{606LGP2}/K^{858RIG-I}, appears to be ordered in a RIG-I RD-RNA complex structures (PDB: 3NCU), extending β -sheet 5 (Fig. 34 A, B) (Wang *et al.* 2010). This also suggests that this region is disordered and flexible in an unbound state and only becomes structured upon ligand interaction of RDs. The extended β -sheet 5 seems to stack on the blunt RNA end in the described lid-like manner.

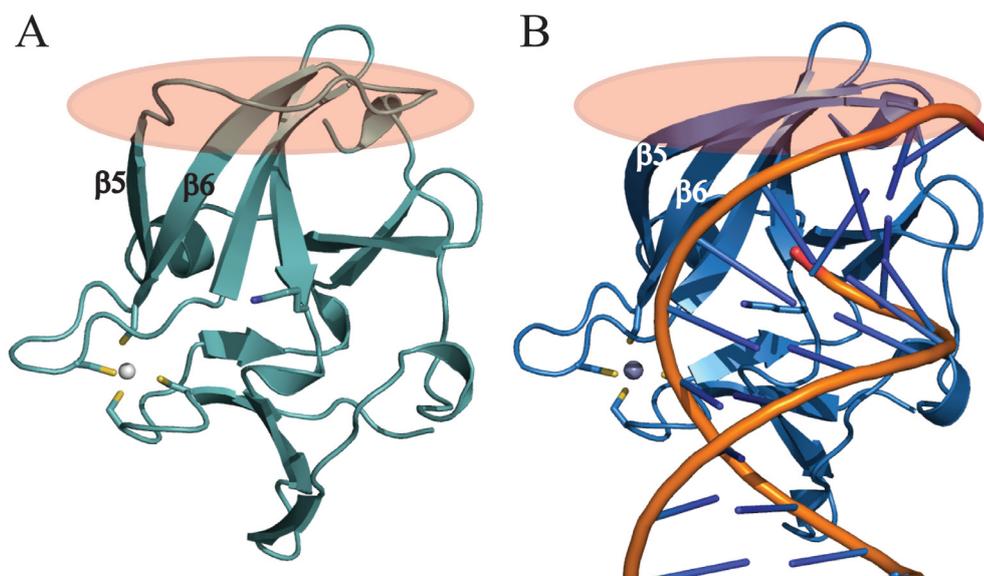


Figure 34 Comparison of the RIG-I RD apo structure (2QFD) (A) with the synthetic 5'-triphosphate bound state (3NCU) (B). β -sheets 5 and 6 and the proposed lid with $\beta 5$ closed on top of the dsRNA ligand are highlighted.

The structure of this RIG-I RD complex with a synthetic 5'-triphosphate blunt end dsRNA (3NCU) has been solved only recently and is also in perfect agreement with the core RNA binding and specificity site model proposed here (Wang *et al.* 2010). Generally, the ability of either LGP2 or RIG-I RDs to bind to the ends of dsRNA would appear to be

advantageous for complex crystallization by allowing a good crystal packing arrangement with two molecules bound to each end of a fairly short blunt end dsRNA (6 or 12 bp).

Dimer formation of LGP2 RD upon dsRNA binding was never observed. This is also confirmed by the crystal structure (PDB: 3EQT) of the complex (Li *et al.* 2009b). Full length LGP2 has however been shown to form dimers, suggesting an activation mechanism related to that of RIG-I (Cui *et al.* 2008; Murali *et al.* 2008).

5.2. LGP2 as a Regulator of RIG-I and MDA5 Signaling

The function and mechanism of LGP2 as a regulator in antiviral innate immune response is puzzling. Previous analysis of LGP2 deficient mice had uncovered a surprising dichotomy of LGP2 function with respect to regulation of RIG-I and MDA5 activity. Whereas LGP2-deficient mice showed reduced susceptibility to viruses recognized by RIG-I, these mice show enhanced susceptibility to challenge with MDA5-specific viruses like EMCV (Venkataraman *et al.* 2007).

To test the LGP2 RD characterizing findings that were gained by *in vitro* analyses and to draw conclusions that could be also applied to full length LGP2, interferon- β reporter assays in HEK293 cells were conducted in collaboration (Pippig *et al.* 2009). Thus, the regulatory effect of LGP2 and the respective contribution of its domains on MDA5 and RIG-I signaling could be assessed (Fig. 35).

To clarify the mechanisms by which LGP2 differentially regulates RIG-I and MDA5 and the role of its RD in these processes, increasing amounts of LGP2 were co-expressed with RIG-I or MDA5 in HEK293 cells. Activation of a luciferase-based interferon- β reporter assay system after stimulation with the appropriate RNA ligand was monitored.

As demonstrated before, when co-expressing LGP2 together with RIG-I, loss of interferon-promoter activation was observed proportional to the level of LGP2 over-expression (Rothenfusser *et al.* 2005; Venkataraman *et al.* 2007). Contrary, when MDA5 was co-expressed with different amounts of LGP2 in the presence of the dsRNA analog poly(I:C) optimal activation, rather than repression was observed at LGP2 expression levels similar to those of MDA5. This was consistent with the weak virus induced MDA5 response seen in LGP2 deficient mice (Venkataraman *et al.* 2007).

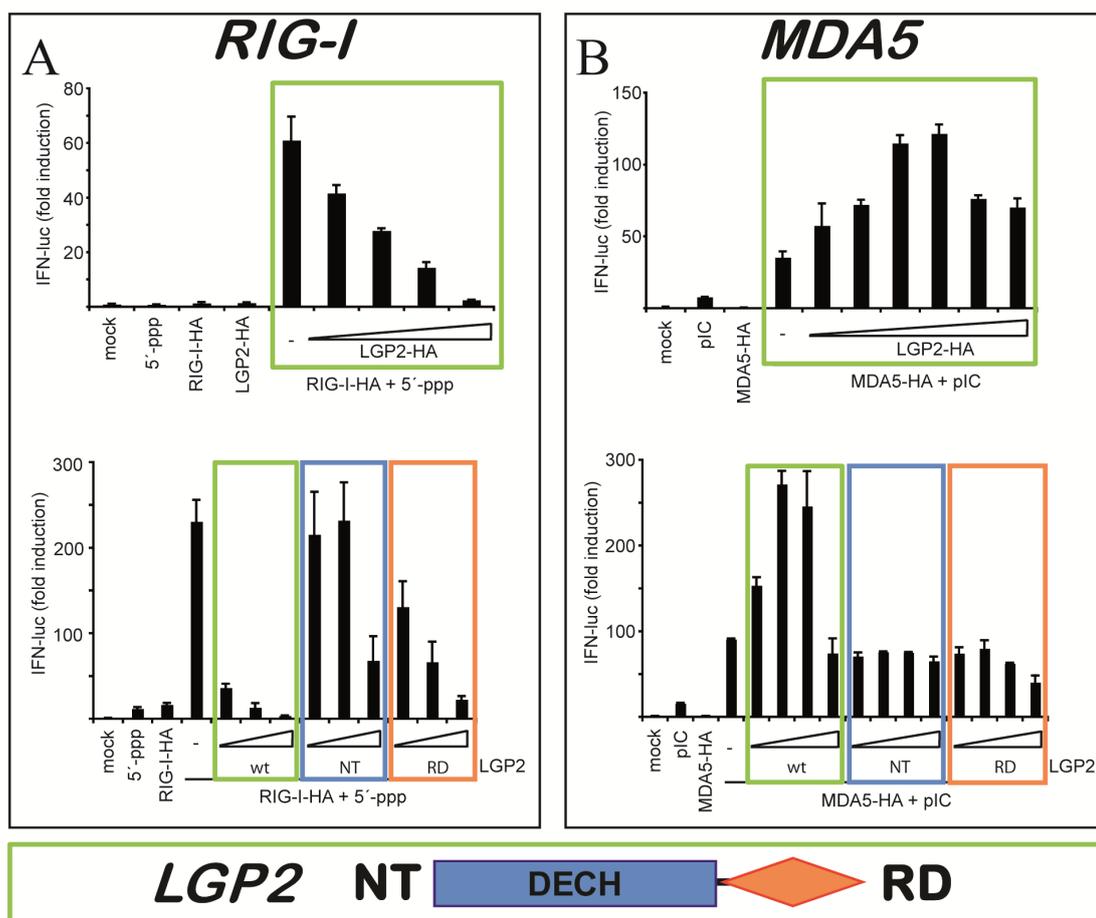


Figure 35 (A) Interferon- β reporter assay of Hek293 cells, transfected with plasmids encoding RIG-I-HA (10 ng plasmid, HA: Hemagglutinin-tag) and different amounts of LGP2-HA (2, 5, 10 and 50 ng plasmid) and stimulated with 5'ppp RNA. Alternatively, cells were transfected with RIG-I-HA and different amounts of wildtype (wt), N-terminal (NT, aa 1-468) or C-terminal (RD, aa 476-678) constructs of LGP2-HA. (B) Cells were transfected with plasmids encoding MDA5-HA (1 ng plasmid) and different amounts of LGP2-HA (0.05, 0.1, 1, 5, 50 and 100 ng plasmid) and stimulated with poly(I:C). Alternatively, cells were transfected with MDA5-HA and different amounts of wt, NT or RD constructs of LGP2-HA.

Relative contributions of the two LGP2 functional domains (DEXH helicase domain - NT and regulatory domain - RD) towards RIG-I repression and MDA5 activation were examined. This was conducted in the presence of different levels of either full-length LGP2 or one of the two functional domains. As has been observed before, LGP2 RD alone is also able to mediate RIG-I repression albeit more weakly than the full length protein (Saito *et al.* 2007). Consistently, an additive contribution of the helicase domain to LGP2-mediated repression of RIG-I is detected (Fig. 35 A).

On the other hand, while full length LGP2-mediated MDA5 activation happened as expected, no stimulation was observed in presence of different amounts of either LGP2 helicase domain or RD only (Fig. 35 B). Hence a fully functional LGP2 is required to mediate a synergistic effect on MDA5, while the inhibitory effect on RIG-I signaling appears additive concerning domain contribution.

Explanations for this controversial role of LGP2 remain speculative. More recently LGP2 has again been demonstrated to exhibit a positive effect on MDA5 but also on RIG-I signaling (Sato *et al.* 2010). Rather than direct interaction a function for LGP2 upstream of RIG-I and MDA5 signaling cascades has thereby been proposed. It can be speculated that LGP2 can bind to RNA in the cytosol that is highly structured or coated by proteins with high affinity. By means of LGP2's helicase and ATPase activity the RNA ligand could then be cleared to allow for accessibility by the other two RLRs.

Consequently, LGP2 can be assumed to act in mechanistically differing ways and various regulatory scenarios could be possible (Fig. 36).

5.3. Possible Mechanisms of LGP2 Interference with MDA5 and RIG-I Signaling

Inhibitory effects of LGP2 have so far only been indicated towards RIG-I signaling. LGP2 RD has been suggested to interact in trans with the RIG-I CARDS. Thereby, RIG-I is kept in an inactive conformation, by preventing CARD exposure for downstream signaling (Fig. 36 A) (Saito *et al.* 2007).

LGP2 has also been demonstrated to interact with the RIG-I downstream signaling mediator IPS-1. It was shown to compete for IKK ϵ binding in a section spanning IPS-1's transmembrane domain and part of its unstructured region. LGP2's binding to IPS-1 is supposedly virus and RNA-independent. Also, IPS-1 RIG-I interaction has been shown to be unaffected. So, this represents a passive mode of inhibition of RIG-I signaling by LGP2 (Fig. 36 B) (Komuro *et al.* 2006; Vitour *et al.* 2007).

A third explanation for LGP2's repressive effect on RIG-I signaling is more in coherence with the finding that both the helicase domain and RD of LGP2 can be inhibitory with effects being additive. In this case the high affinity of LGP2 and its domains to potential RIG-I RNA-ligands would account for its functioning as a sink for dsRNA in the cell. Thereby RNA is inaccessible for recognition by RIG-I. This latter scenario would presumably be dosage dependent, LGP2 might confer regulation in a negative feedback loop when upregulated in response to viral stimuli.

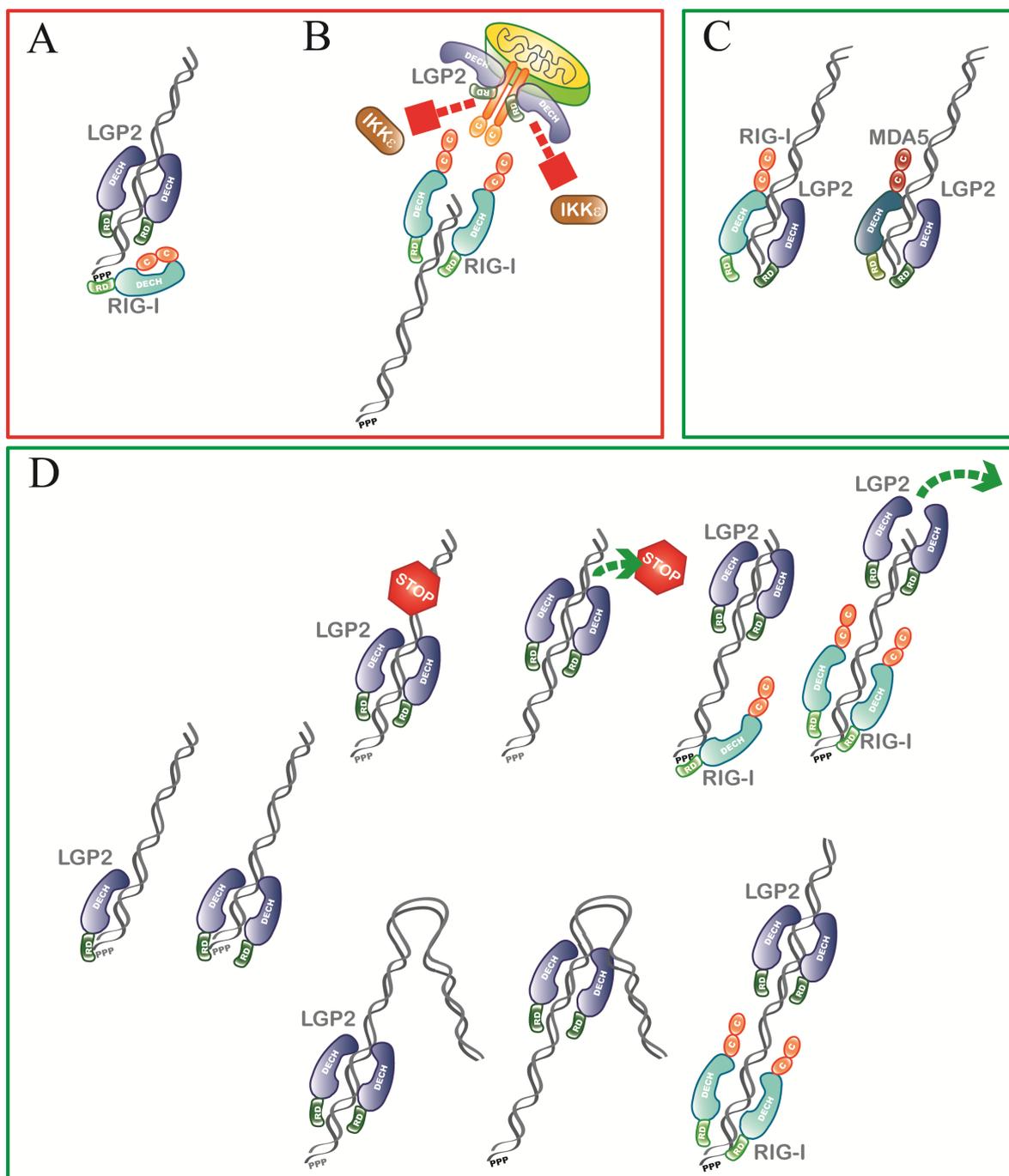


Figure 36 Possible mechanisms of RIG-I and MDA5 regulation by LGP2. Red and green frames represent inhibitory or activating effects, respectively. (A) RIG-I inhibition by trans-interaction of LGP2 RD with its CARDS. (B) Passive inhibition by LGP2 competition with IKKε for IPS-1 binding. (C) Possibly active heterodimeric RLR-RNA complexes and (D) LGP2's mode of dsRNA interaction and possible processing by removal of bound or stalled protein or secondary structure. Domains are depicted in shades of green (RDs), red (CARDS – “C”) and blue (helicase domain – “DECH”).

A further explanation is that LGP2 could be an activator of RIG-I and MDA5 signaling by forming heterodimers upon RNA exposure that are capable of signal transduction. This especially applies to MDA5, whose exact RNA PAMP is still subject to controversial discussion, and where heterodimer formation might be an explanation for missing knowledge about an MDA5-only ligand (Fig. 36 C). Even though, heterotypic interactions between the RLRs have been shown by whole cell extract pulldowns of virus infected cells, so far no evidence for such complexes exist from *in vitro* experiments (Komuro *et al.* 2006).

Finally, recent studies indicate an activating role of LGP2 operating upstream of RIG-I and MDA5 (Sato *et al.* 2010). LGP2 might interact with RNA similarly to RIG-I by first docking to dsRNA with its RD and subsequent dimerization of the helicase domain. It is possibly more versatile in binding to various RNA conformations and could therefore make initially unrecognized RNA accessible to RIG-I and MDA5. By helicase movement along the dsRNA strands LGP2 could remove other RNA bound or stalled proteins or eliminate secondary structure. Thus translocation of either MDA5 or RIG-I would be enabled subsequently (Fig. 36 D). If the helicase activity of RIG-I and MDA5 was actually required for their signaling, LGP2 could also function in releasing stalled RLRs, thus ensuring their availability for further signal transduction.

LGP2 could also represent a means of bringing sensor (RIG-I or MDA5) and signaling or mediator molecules (like IPS-1) in close proximity to stimulate downstream signaling. If this was the case, the observed LGP2 interaction with IPS-1 would appear in an activating context.

In general, LGP2's regulatory effect on RLR signaling is still controversial. It is likely to be dosage dependent and highly regulated itself. Hence, a stimulating effect on RIG-I signaling as well as its inhibition by LGP2 are not mutually exclusive.

At low or normal expression levels LGP2 could activate RIG-I and MDA5 signaling by different possible mechanisms (Fig. 36 C, D).

During progress of the inflammation and to prevent exaggerated immune reactions or false-positive signaling by RIG-I, LGP2 might be upregulated. High levels of LGP2 would then increase the competition for RNA in the cytosol. LGP2 can bind to RNA with high affinity and less specificity than RIG-I. This prevents RIG-I's binding to RNA and suppresses subsequent signaling events (Fig. 36 A). This interpretation is also in coherence with the finding that LGP2 RD by itself is able to abrogate RIG-I activity *in vivo*. LGP2 RD could bind to 5'-triphosphate containing dsRNA ends and block them, thus impeding RIG-I loading onto the RNA.

As yet unidentified proteins might further be involved in this tightly regulated network that could aid LGP2's modulating effect.

It seems that the variability of results, gained by *in vivo* assays which demonstrate LGP2's role as regulator of RIG-I and MDA5 signaling, might partially originate from the use of varying stimuli. It can be assumed that synthetic RNA ligands and especially Poly(I:C) do not cause an entirely physiological relevant response. Moreover, RIG-I and MDA5 exhibit susceptibility towards different viruses. Thus, a careful selection of the viral stimulus as well as a greater knowledge of structural characteristics of the RNA it harbors, are indispensable to allow for proper conclusions from the assay results.

5.4. The AIM2 Inflammasome – Preliminary Functional Insights

RLRs are capable of directly triggering an immune response upon sensing pathogenic RNA in the cytosol. The AIM2 inflammasome and others however, represent a second stage of inflammatory response by processing inflammatory cytokines like pro-IL1 β that result from the primary reaction.

The AIM2 inflammasome is unique compared to other inflammasomes and yet more unknown inflammasomes might exist in this tightly regulated signaling network. Even RIG-I that has recently been shown to interact with the common inflammasome adaptor ASC, as does AIM2, might be a candidate for a new RNA responsive inflammasome-like platform.

Recombinantly expressed AIM2 could be shown to bind dsDNA and certain crystallization suited DNA-ligands have been identified.

It is proposed that AIM2 forms higher oligomeric complexes upon binding to DNA, whereby oligomerization is mediated by the conserved MFHATVAT motif located in the C-terminal DNA binding HIN domain. Yet, in binding studies a maximum of only two binding sites are apparent, judging from two distinctly shifted bands in EMSAs. However the longest used oligonucleotide had 35 bp only and it has been shown that interaction is length dependent and at least around 40 bp are required to trigger AIM2 inflammasome formation *in vivo* (Fernandes-Alnemri *et al.* 2009; Roberts *et al.* 2009). This indicates that AIM2 oligomerization is possibly facilitated by binding of a certain number of AIM2 molecules next to each other on a dsDNA strand.

Binding per se seems to be rather strong even with shorter dsDNA. Also protein-DNA complexes are stable enough to withstand purification in size exclusion chromatography. Aggregation of the human AIM2-HIN domain constructs when exposed to dsDNA might also indicate the formation of larger complexes.

An interaction between recombinant AIM2 and ASC or ASC-PYD in presence of DNA could be demonstrated. This represents a first step towards identifying crystallizable bi- or tripartite inflammasomal subcomplexes. Future improvement of the stability of the individual AIM2 inflammasome components or determination of a self-stabilizing complex will however be required for this purpose. While AIM2 was rarely prone to aggregation, it undergoes rapid N-terminal (PYD domain) degradation. ASC and its PYD-only constructs proved challenging to handle, because of the strong bipolar character that would often lead to aggregation of the purified protein after GST-tag cleavage, especially in physiological pH ranges.

For the NLRP1 inflammasome a penta- or heptameric ring-like structure has been proposed (Faustin *et al.* 2007) and it is also possible that AIM2 would form ordered

multiprotein platforms upon recognition of cytosolic dsDNA. If the AIM2 inflammasome was also appearing in heptameric clusters, either a ring-like shape or a more linear conformation could be possible (Fig. 37), however no experimental evidence exists for this. The AIM2 inflammasome could however also be more of an unordered cluster induced by multiple neighboring AIM2 binding sites on dsDNA. In the future, it will be important to determine the exact inflammasome stoichiometry. It is also of interest to see whether in such complexes each AIM2 molecule would exhibit DNA contacts or whether binding of a sub-complex is enough to trigger further building block recruitment. In this model oligomerization could be mediated by either AIM2's dimerization motive or by ASC – ASC interactions.

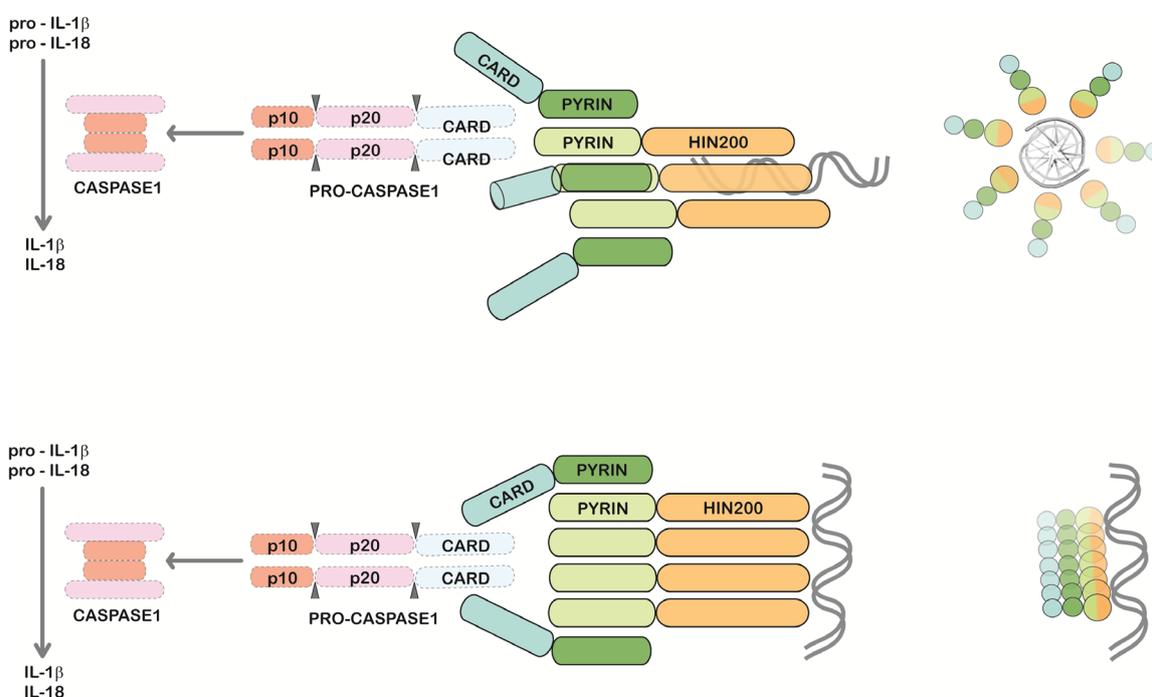


Figure 37 Theoretical schemes for complex formation of the AIM2 inflammasome. Stoichiometries are indicative only. AIM2 binds dsDNA with its HIN domain and attaches ASC *via* homotypic PYD interaction. Subsequently pro-caspase 1 is recruited by CARD interaction to form a catalytically active caspase complex after auto-cleavage.

5.4.1. Crystallization of AIM2 – DNA Complexes

Even though crystals that possibly contain the AIM2-DNA complex (or HIN-DNA) could be already grown, improvement is required to gain good quality crystals for final structure determination. Major problems were a generally high number of crystallization nuclei and the crystals growing bound to the plate. Crystallization in an agarose gel matrix might help to circumvent this. The so far used DNA-hairpins may also not be ideal for optimal crystal packing and more DNA species, namely blunt end dsDNA of different lengths, need to be tested for crystallization. Once a stable complex of AIM2 and ASC/ASC-PYD is established in large scale this should be crystallized in presence of DNA as well. Despite the proposed higher oligomeric molecular structure of the inflammasome and judging from here shown results, it appears likely that certain AIM2 inflammasome subcomplexes could crystallize.

5.5. Comparison of RLRs and AIM2

RLR and AIM2 have several things in common. They are cytosolic receptors directed against pathogenic nucleic acids that enter the cell during infection. Upon recognition of their respective PAMP, they trigger pro-inflammatory or antiviral responses. So far, RIG-I is the best characterized RLR and believed to form dimers after interaction with 5'-triphosphate dsRNA, that can also originate from RNA Polymerase III transcribed cytosolic DNA. RIG-I signals downstream *via* its CARD domains. The RIG-I pathway gets thereby localized to the mitochondrial membrane, forming a large antiviral signaling focus.

AIM2 on the other hand forms higher oligomeric assemblies upon interaction with dsDNA that is marked as foreign by its occurrence in the cytosol. For ultimate activation, AIM2 binds the inflammasomal adaptor ASC that then recruits pro-caspase 1 also *via* CARD domain interaction. While RLRs MDA5 and RIG-I mainly trigger the production of interferons and pro-interleukins, the fully functional AIM2 inflammasome processes the latter. This again represents a common intersection between the two pathways.

Finally, RIG-I has recently been shown to also interact with ASC through CARD domain interaction. So, it can be speculated that it is able to form an inflammasome-like structure similar to AIM2, but in response to RNA rather than DNA.

LGP2 is different due to its lacking an active signaling function. While RIG-I and MDA5 have CARDs and AIM2 has a Pyrin domain for downstream signal transduction, LGP2 is restricted to a regulatory role.

In the AIM2 inflammasome network *mp202*, or similarly acting possible splice variants of human AIM2, come closest to be comparable with LGP2. They represent an analog way of regulation by being able to interact with dsDNA and AIM2, but lacking the adaptor PYD required for referring a signal to ASC.

Generally, it seems to be a common regulatory motive in PRRs of the innate immune system to have a highly similar counterpart that is deficient in a domain required for downstream interaction. The inhibitory or regulatory effect then lies in the association of this limited counterpart with the fully functional PRR. Other examples for this are COPs and POPs that either harbor only a Pyrin domain or CARD and are thus capable of interfering with ASC signaling.

6. *Summary*

With LGP2, a member of the RIG-I-like receptor family and AIM2, a recently discovered, unusual inflammasome module, two cytosolic pattern recognition receptors of the innate immune system could be characterized in the course of this PhD thesis.

LGP2 has been shown to be a potent and specific receptor of viral, especially duplex RNA, despite its lack of CARD domains that are thought to be indispensable for direct signal relaying in the RLR pathway. Hence, LGP2 is not directly mediating the expression of pro-inflammatory cytokines, but it is involved in the regulation of RLR signaling and thus cytokine production.

AIM2 is only partially capable of an upregulation of cytokine expression levels upon infection. It is more importantly required for the processing of cytokine precursors originating for example from an active RLR pathway. In contrast to the RNA receptor LGP2, AIM2 has been shown to specifically bind dsDNA in the cytosol.

Together, LGP2 and RLRs in general as well as AIM2 represent crucial entities directed against diverse pathogens that can infect a cell, by means of detecting their nucleic acid constituents.

It could be demonstrated that the regulatory C-terminal domain in RLRs is indispensable for their RNA recognition and interaction. The structure of LGP2 RD could be solved and it has been shown to exhibit a strong affinity towards blunt dsRNA. Notably LGP2 RD facilitates this contact entirely independent of the presence of phosphates at the 3'- or 5'- RNA-ends. This is in direct contrast to RIG-I that exhibits strongest binding and activation in presence of a 5'-triphosphate moiety, a common feature in viral genomes that usually remain unprocessed or non-capped. Although the RD of LGP2 exhibits high structural similarity to the RD of RIG-I, it exhibits remarkably different RNA binding specificities. Thus, the results imply that the regulatory domains of RLR confer their pattern specificity.

The structure and data suggest that RDs of all three RIG-I like helicases are highly conserved RNA binding elements with a common core RNA binding site but specific adaptation to their respectively recognized patterns. A shared RNA binding and varying specificity sites in RLR RDs were postulated. In particular the properties found specific for LGP2 gave rise to possible explanations of the regulatory mechanism it exhibits on RIG-I and MDA5 signaling. In case of an inhibitory function of LGP2 on RLR signaling, this might occur as a negative feedback to prevent continuous stimulation of pro-inflammatory factors. Either the sequestration of RNA by high affinity exhibiting LGP2 or the formation of a

hetero-complex of LGP2 with RIG-I or MDA5 that is not capable of signaling are possibilities. The latter case has however never been shown and could still be functional in signal transduction to IPS-1. For the activating function of LGP2 and aside from the possible formation of active hetero-dimers, a role upstream of RIG-I and MDA5 has been suggested. LGP2 would thereby act as a rather universal RNA receptor capable of processing and clearing viral RNA structures to make them accessible for RIG-I and MDA5. The postulated role of LGP2 RD could account for both the inhibitory and activating task and it seems that mainly the dose of LGP2 is decisive as for which effect prevails.

The AIM2 inflammasome project is still emerging. A structure of AIM2 or its complex with DNA could not be solved. However, progress has been made towards the identification of possible DNA ligands that are favorable to result in diffracting crystals harboring an AIM2-DNA complex in the future. This is especially interesting since inflammasome formation *in vivo* has been only reported to occur with dsDNAs containing at least 40 base pairs. *In vitro*, however, strong complex formation is already achieved with 11 bp, which should allow for good crystal packing in co-crystallization experiments. Crystals of a complex containing the HIN domain of AIM2 or a supposedly N-terminally degraded full version and short DNA hairpins were able to be reproducibly grown.

The identification of recombinant AIM2 and ASC constructs capable to interact with each other represents a huge step towards the *in vitro* reconstitution of the AIM2 inflammasome or subcomplexes of it. In the future this should provide insight into the stoichiometry and complex formation of this pro-inflammatory platform.

The AIM2 inflammasome, containing AIM2, ASC as well as caspase-1 in unknown stoichiometry, tends to form into huge aggregates in cells infected by pathogenic DNA. Nevertheless, judging from the obtained results, it appears as a reasonable target for crystallization of at least subcomplexes, for example the DNA-AIM2-ASC or DNA-AIM2-PYD(ASC 1-93) part.

In conclusion, good progress has been made in testing AIM2's DNA binding capacity *in vitro* and identifying DNA ligands suited for co-crystallization. Generally, advancement has been made into finding constructs that might readily crystallize. Hence, the solution of high-resolution structures of parts of the AIM2 inflammasome in the near future seems realistic.

To the RLR field this PhD project is of great importance. It especially helped to extend the understanding of LGP2's role in RLR signaling regulation. Also, insight into the mechanism underlying RNA ligand binding and the varying specificity apparent in RLRs has been provided.

7. References

- Abdul-Sater, A. A., Said-Sadier, N., Ojcius, D. M., Yilmaz, O. and Kelly, K. A. (2009). "Inflammasomes bridge signaling between pathogen identification and the immune response." **Drugs Today (Barc)** 45 Suppl B: 105-112.
- Ablasser, A., Bauernfeind, F., Hartmann, G., Latz, E., Fitzgerald, K. A. and Hornung, V. (2009). "RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate." **Nat Immunol** 10(10): 1065-1072.
- Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. and Terwilliger, T. C. (2004). "Recent developments in the PHENIX software for automated crystallographic structure determination." **J Synchrotron Radiat** 11(Pt 1): 53-55.
- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. and Terwilliger, T. C. (2002). "PHENIX: building new software for automated crystallographic structure determination." **Acta Crystallogr D Biol Crystallogr** 58(Pt 11): 1948-1954.
- Alberts, B., Johnson, A., Walter, P., Raff, M., Roberts, K. and Lewis, J. (2002). "Molecular Biology of the Cell", Taylor & Francis.
- Albrecht, M., Choubey, D. and Lengauer, T. (2005). "The HIN domain of IFI-200 proteins consists of two OB folds." **Biochem Biophys Res Commun** 327(3): 679-687.
- Arnaud-Barbe, N., Cheynet-Sauvion, V., Oriol, G., Mandrand, B. and Mallet, F. (1998). "Transcription of RNA templates by T7 RNA polymerase." **Nucleic Acids Res** 26(15): 3550-3554.
- Barton, G. M. (2008). "A calculated response: control of inflammation by the innate immune system." **J Clin Invest** 118(2): 413-420.
- Berger, I., Fitzgerald, D. J. and Richmond, T. J. (2004). "Baculovirus expression system for heterologous multiprotein complexes." **Nat Biotechnol** 22(12): 1583-1587.
- Bertin, J. and DiStefano, P. S. (2000). "The PYRIN domain: a novel motif found in apoptosis and inflammation proteins." **Cell Death Differ** 7(12): 1273-1274.
- Bonilla, F. A. and Oettgen, H. C. (2010). "Adaptive immunity." **J Allergy Clin Immunol** 125(2 Suppl 2): S33-40.
- Brunger, A. T. (2007). "Version 1.2 of the Crystallography and NMR system." **Nat Protoc** 2(11): 2728-2733.
- Burckstummer, T., Baumann, C., Bluml, S., Dixit, E., Durnberger, G., Jahn, H., Planyavsky, M., Bilban, M., Colinge, J., Bennett, K. L. and Superti-Furga, G. (2009). "An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome." **Nat Immunol** 10(3): 266-272.
- Cazenave, C. and Uhlenbeck, O. C. (1994). "RNA template-directed RNA synthesis by T7 RNA polymerase." **Proc Natl Acad Sci U S A** 91(15): 6972-6976.

- Chaplin, D. D. (2010). "Overview of the immune response." **J Allergy Clin Immunol** 125(2 Suppl 2): S3-23.
- Cheng, J., Waite, A. L., Tkaczyk, E. R., Ke, K., Richards, N., Hunt, A. J. and Gumucio, D. L. (2010). "Kinetic properties of ASC protein aggregation in epithelial cells." **J Cell Physiol** 222(3): 738-747.
- Chiu, Y. H., Macmillan, J. B. and Chen, Z. J. (2009). "RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway." **Cell** 138(3): 576-591.
- Choubey, D. and Lengyel, P. (1995). "Binding of an interferon-inducible protein (p202) to the retinoblastoma protein." **J Biol Chem** 270(11): 6134-6140.
- Choubey, D. and Panchanathan, R. (2008). "Interferon-inducible Ifi200-family genes in systemic lupus erythematosus." **Immunol Lett** 119(1-2): 32-41.
- Choubey, D., Walter, S., Geng, Y. and Xin, H. (2000). "Cytoplasmic localization of the interferon-inducible protein that is encoded by the AIM2 (absent in melanoma) gene from the 200-gene family." **FEBS Lett** 474(1): 38-42.
- Chuang, T. H., Lee, J., Kline, L., Mathison, J. C. and Ulevitch, R. J. (2002). "Toll-like receptor 9 mediates CpG-DNA signaling." **J Leukoc Biol** 71(3): 538-544.
- Cresswell, K. S., Clarke, C. J., Jackson, J. T., Darcy, P. K., Trapani, J. A. and Johnstone, R. W. (2005). "Biochemical and growth regulatory activities of the HIN-200 family member and putative tumor suppressor protein, AIM2." **Biochem Biophys Res Commun** 326(2): 417-424.
- Cui, S., Eisenacher, K., Kirchhofer, A., Brzozka, 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." **Mol Cell** 29(2): 169-179.
- Dawson, M. J. and Trapani, J. A. (1996). "HIN-200: a novel family of IFN-inducible nuclear proteins expressed in leukocytes." **J Leukoc Biol** 60(3): 310-316.
- de Alba, E. (2009). "Structure and interdomain dynamics of apoptosis-associated speck-like protein containing a CARD (ASC)." **J Biol Chem** 284(47): 32932-32941.
- Ding, Y., Wang, L., Su, L. K., Frey, J. A., Shao, R., Hunt, K. K. and Yan, D. H. (2004). "Antitumor activity of IFIX, a novel interferon-inducible HIN-200 gene, in breast cancer." **Oncogene** 23(26): 4556-4566.
- Dodson, E. (2003). "Is it jolly SAD?" **Acta Crystallogr D Biol Crystallogr** 59(Pt 11): 1958-1965.
- Dunkelberger, J. R. and Song, W. C. (2010). "Complement and its role in innate and adaptive immune responses." **Cell Res** 20(1): 34-50.
- Eisenbarth, S. C., Colegio, O. R., O'Connor, W., Sutterwala, F. S. and Flavell, R. A. (2008). "Crucial role for the Nalp3 inflammasome in the immunostimulatory properties of aluminium adjuvants." **Nature** 453(7198): 1122-1126.
- Emsley, P. and Cowtan, K. (2004). "Coot: model-building tools for molecular graphics." **Acta Crystallogr D Biol Crystallogr** 60(Pt 12 Pt 1): 2126-2132.
- Fairbrother, W. J., Gordon, N. C., Humke, E. W., O'Rourke, K. M., Starovasnik, M. A., Yin, J. P. and Dixit, V. M. (2001). "The PYRIN domain: a member of the death domain-fold superfamily." **Protein Sci** 10(9): 1911-1918.

- Faustin, B., Lartigue, L., Bruey, J. M., Luciano, F., Sergienko, E., Bailly-Maitre, B., Volkman, N., Hanein, D., Rouiller, I. and Reed, J. C. (2007). "Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase-1 activation." **Mol Cell** 25(5): 713-724.
- Fernandes-Alnemri, T., Wu, J., Yu, J. W., Datta, P., Miller, B., Jankowski, W., Rosenberg, S., Zhang, J. and Alnemri, E. S. (2007). "The pyroptosome: a supramolecular assembly of ASC dimers mediating inflammatory cell death via caspase-1 activation." **Cell Death Differ** 14(9): 1590-1604.
- 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(7237): 509-513.
- Ferrero-Miliani, L., Nielsen, O. H., Andersen, P. S. and Girardin, S. E. (2007). "Chronic inflammation: importance of NOD2 and NALP3 in interleukin-1beta generation." **Clin Exp Immunol** 147(2): 227-235.
- Gallucci, S. and Matzinger, P. (2001). "Danger signals: SOS to the immune system." **Curr Opin Immunol** 13(1): 114-119.
- Gattin, Z. and van Gunsteren, W. F. (2008). "A molecular dynamics study of the ASC and NALP1 pyrin domains at neutral and low pH." **ChemBiochem** 9(6): 923-933.
- Geddes, B. J., Wang, L., Huang, W. J., Lavellee, M., Manji, G. A., Brown, M., Jurman, M., Cao, J., Morgenstern, J., Merriam, S., Glucksmann, M. A., DiStefano, P. S. and Bertin, J. (2001). "Human CARD12 is a novel CED4/Apaf-1 family member that induces apoptosis." **Biochem Biophys Res Commun** 284(1): 77-82.
- Hartmann, R. K., Bindereif, A., Schön, A. and Westhof, E. (2005). "Handbook of RNA biochemistry."
- Hendrickson, W. A., Horton, J. R. and LeMaster, D. M. (1990). "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure." **EMBO J** 9(5): 1665-1672.
- Holm, L. and Park, J. (2000). "DaliLite workbench for protein structure comparison." **Bioinformatics** 16(6): 566-567.
- 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(7237): 514-518.
- Hornung, V., Bauernfeind, F., Halle, A., Samstad, E. O., Kono, H., Rock, K. L., Fitzgerald, K. A. and Latz, E. (2008). "Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization." **Nat Immunol** 9(8): 847-856.
- Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K. K., Schlee, M., Endres, S. and Hartmann, G. (2006). "5'-Triphosphate RNA is the ligand for RIG-I." **Science** 314(5801): 994-997.
- Ippagunta, S. K., Brand, D. D., Luo, J., Boyd, K. L., Calabrese, C., Stienstra, R., Van de Veerdonk, F. L., Netea, M. G., Joosten, L. A., Lamkanfi, M. and Kanneganti, T. D. (2010). "Inflammasome-independent role of apoptosis-associated speck-like protein containing a CARD (ASC) in T cell priming is critical for collagen-induced arthritis." **J Biol Chem** 285(16): 12454-12462.
- Johnstone, R. W., Wei, W., Greenway, A. and Trapani, J. A. (2000). "Functional interaction between p53 and the interferon-inducible nucleoprotein IFI 16." **Oncogene** 19(52): 6033-6042.

- Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." **J Appl Cryst** 26: 795-800.
- Kato, H., Sato, S., Yoneyama, M., Yamamoto, M., Uematsu, S., Matsui, K., Tsujimura, T., Takeda, K., Fujita, T., Takeuchi, O. and Akira, S. (2005). "Cell type-specific involvement of RIG-I in antiviral response." **Immunity** 23(1): 19-28.
- 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." **J Exp Med** 205(7): 1601-1610.
- Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K. J., Yamaguchi, O., Otsu, K., Tsujimura, T., Koh, C. S., Reis e Sousa, C., Matsuura, Y., Fujita, T. and Akira, S. (2006). "Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses." **Nature** 441(7089): 101-105.
- Komuro, A. and Horvath, C. M. (2006). "RNA- and virus-independent inhibition of antiviral signaling by RNA helicase LGP2." **J Virol** 80(24): 12332-12342.
- Kumagai, Y. and Akira, S. (2010). "Identification and functions of pattern-recognition receptors." **J Allergy Clin Immunol** 125(5): 985-992.
- Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." **Nature** 227(5259): 680-685.
- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N. (2005). "ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures." **Nucleic Acids Res** 33(Web Server issue): W299-302.
- Landolfo, S., Gariglio, M., Gribaudo, G. and Lembo, D. (1998). "The Ifi 200 genes: an emerging family of IFN-inducible genes." **Biochimie** 80(8-9): 721-728.
- Latz, E. (2010). "The inflammasomes: mechanisms of activation and function." **Curr Opin Immunol** 22(1): 28-33.
- Lengyel, P. and Liu, C. J. (2010). "The p200 family protein p204 as a modulator of cell proliferation and differentiation: a brief survey." **Cell Mol Life Sci** 67(3): 335-340.
- Li, X., Lu, C., Stewart, M., Xu, H., Strong, R. K., Igumenova, T. and Li, P. (2009a). "Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5." **Arch Biochem Biophys** 488(1): 23-33.
- Li, X., Ranjith-Kumar, C. T., Brooks, M. T., Dharmaiah, S., Herr, A. B., Kao, C. and Li, P. (2009b). "The RIG-I-like receptor LGP2 recognizes the termini of double-stranded RNA." **J Biol Chem** 284(20): 13881-13891.
- Loo, Y. M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira, S., Gill, M. A., Garcia-Sastre, A., Katze, M. G. and Gale, M., Jr. (2008). "Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity." **J Virol** 82(1): 335-345.
- Ludlow, L. E., Johnstone, R. W. and Clarke, C. J. (2005). "The HIN-200 family: more than interferon-inducible genes?" **Exp Cell Res** 308(1): 1-17.

- Malathi, K., Dong, B., Gale, M., Jr. and Silverman, R. H. (2007). "Small self-RNA generated by RNase L amplifies antiviral innate immunity." **Nature** 448(7155): 816-819.
- Martinon, F., Burns, K. and Tschopp, J. (2002). "The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta." **Mol Cell** 10(2): 417-426.
- Martinon, F., Hofmann, K. and Tschopp, J. (2001). "The pyrin domain: a possible member of the death domain-fold family implicated in apoptosis and inflammation." **Curr Biol** 11(4): R118-120.
- Masumoto, J., Taniguchi, S., Ayukawa, K., Sarvotham, H., Kishino, T., Niikawa, N., Hidaka, E., Katsuyama, T., Higuchi, T. and Sagara, J. (1999). "ASC, a novel 22-kDa protein, aggregates during apoptosis of human promyelocytic leukemia HL-60 cells." **J Biol Chem** 274(48): 33835-33838.
- Masumoto, J., Taniguchi, S. and Sagara, J. (2001). "Pyrin N-terminal homology domain- and caspase recruitment domain-dependent oligomerization of ASC." **Biochem Biophys Res Commun** 280(3): 652-655.
- McCoy, A. J., Grosse-Kunstleve, P. D., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007a). "Phaser crystallographic software." **J Appl Cryst** 40: 658-674.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007b). "Phaser crystallographic software." **J Appl Crystallogr** 40(Pt 4): 658-674.
- Min, W., Ghosh, S. and Lengyel, P. (1996). "The interferon-inducible p202 protein as a modulator of transcription: inhibition of NF-kappa B, c-Fos, and c-Jun activities." **Mol Cell Biol** 16(1): 359-368.
- Moriya, M., Taniguchi, S., Wu, P., Liepinsh, E., Otting, G. and Sagara, J. (2005). "Role of charged and hydrophobic residues in the oligomerization of the PYRIN domain of ASC." **Biochemistry** 44(2): 575-583.
- Murali, A., Li, X., Ranjith-Kumar, C. T., Bhardwaj, K., Holzenburg, A., Li, P. and Kao, C. C. (2008). "Structure and function of LGP2, a DEX(D/H) helicase that regulates the innate immunity response." **J Biol Chem** 283(23): 15825-15833.
- Muruve, D. A., Petrilli, V., Zaiss, A. K., White, L. R., Clark, S. A., Ross, P. J., Parks, R. J. and Tschopp, J. (2008). "The inflammasome recognizes cytosolic microbial and host DNA and triggers an innate immune response." **Nature** 452(7183): 103-107.
- Ohtsuka, T., Ryu, H., Minamishima, Y. A., Macip, S., Sagara, J., Nakayama, K. I., Aaronson, S. A. and Lee, S. W. (2004). "ASC is a Bax adaptor and regulates the p53-Bax mitochondrial apoptosis pathway." **Nat Cell Biol** 6(2): 121-128.
- Pawlowski, K., Pio, F., Chu, Z., Reed, J. C. and Godzik, A. (2001). "PAAD - a new protein domain associated with apoptosis, cancer and autoimmune diseases." **Trends Biochem Sci** 26(2): 85-87.
- Pichlmair, A., Schulz, O., Tan, C. P., Naslund, T. I., Liljestrom, P., Weber, F. and Reis e Sousa, C. (2006). "RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates." **Science** 314(5801): 997-1001.
- Pichlmair, A., Schulz, O., Tan, C. P., Rehwinkel, J., Kato, H., Takeuchi, O., Akira, S., Way, M., Schiavo, G. and Reis e Sousa, C. (2009). "Activation of MDA5 requires higher-order RNA structures generated during virus infection." **J Virol** 83(20): 10761-10769.

- 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 Res** 37(6): 2014-2025.
- Poeck, H., Bscheider, M., Gross, O., Finger, K., Roth, S., Rebsamen, M., Hanneschlager, N., Schlee, M., Rothenfusser, S., Barchet, W., Kato, H., Akira, S., Inoue, S., Endres, S., Peschel, C., Hartmann, G., Hornung, V. and Ruland, J. (2010). "Recognition of RNA virus by RIG-I results in activation of CARD9 and inflammasome signaling for interleukin 1 beta production." **Nat Immunol** 11(1): 63-69.
- Ranaivoson, F. M., Neiers, F., Kauffmann, B., Boschi-Muller, S., Branlant, G. and Favier, F. (2009). "Methionine sulfoxide reductase B displays a high level of flexibility." **J Mol Biol** 394(1): 83-93.
- Read, R. J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." **Acta Crystallogr D Biol Crystallogr** 57(Pt 10): 1373-1382.
- Roberts, T. L., Idris, A., Dunn, J. A., Kelly, G. M., Burnton, C. M., Hodgson, S., Hardy, L. L., Garceau, V., Sweet, M. J., Ross, I. L., Hume, D. A. and Stacey, K. J. (2009). "HIN-200 proteins regulate caspase activation in response to foreign cytoplasmic DNA." **Science** 323(5917): 1057-1060.
- Rothenfusser, S., Goutagny, N., DiPerna, G., Gong, M., Monks, B. G., Schoenemeyer, A., Yamamoto, M., Akira, S. and Fitzgerald, K. A. (2005). "The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I." **J Immunol** 175(8): 5260-5268.
- Saito, T. and Gale, M., Jr. (2008a). "Differential recognition of double-stranded RNA by RIG-I-like receptors in antiviral immunity." **J Exp Med** 205(7): 1523-1527.
- Saito, T., Hirai, R., Loo, Y. M., Owen, D., Johnson, C. L., Sinha, S. C., Akira, S., Fujita, T. and Gale, M., Jr. (2007). "Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2." **Proc Natl Acad Sci U S A** 104(2): 582-587.
- Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J. and Gale, M., Jr. (2008b). "Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA." **Nature** 454(7203): 523-527.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). "Molecular Cloning - A laboratory Manual".
- 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." **Proc Natl Acad Sci U S A** 107(4): 1512-1517.
- Schlee, M., Roth, A., Hornung, V., Hagmann, C. A., Wimmenauer, V., Barchet, W., Coch, C., Janke, M., Mihailovic, A., Wardle, G., Juranek, S., Kato, H., Kawai, T., Poeck, H., Fitzgerald, K. A., Takeuchi, O., Akira, S., Tuschl, T., Latz, E., Ludwig, J. and 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(1): 25-34.
- Schmidt, A., Schwerd, T., Hamm, W., Hellmuth, J. C., Cui, S., Wenzel, M., Hoffmann, F. S., Michallet, M. C., Besch, R., Hopfner, K. P., Endres, S. and Rothenfusser, S. (2009). "5'-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I." **Proc Natl Acad Sci U S A** 106(29): 12067-12072.
- Schroder, K. and Tschopp, J. (2010a). "The inflammasomes." **Cell** 140(6): 821-832.
- Schroder, K., Zhou, R. and Tschopp, J. (2010b). "The NLRP3 inflammasome: a sensor for metabolic danger?" **Science** 327(5963): 296-300.

- Schubert, S., Gul, D. C., Grunert, H. P., Zeichhardt, H., Erdmann, V. A. and Kurreck, J. (2003). "RNA cleaving '10-23' DNazymes with enhanced stability and activity." **Nucleic Acids Res** 31(20): 5982-5992.
- Stehlik, C. and Dorfleutner, A. (2007). "COPs and POPs: modulators of inflammasome activity." **J Immunol** 179(12): 7993-7998.
- Stein, N. (2008). "CHAINSAW: a program for mutating pdb files used as templates in molecular replacement." **Journal of Applied Crystallography** 41 (3): 641-643.
- Stutz, A., Golenbock, D. T. and Latz, E. (2009). "Inflammasomes: too big to miss." **J Clin Invest** 119(12): 3502-3511.
- Suzuki, T., Franchi, L., Toma, C., Ashida, H., Ogawa, M., Yoshikawa, Y., Mimuro, H., Inohara, N., Sasakawa, C. and Nunez, G. (2007). "Differential regulation of caspase-1 activation, pyroptosis, and autophagy via Ipaf and ASC in Shigella-infected macrophages." **PLoS Pathog** 3(8): e111.
- Takahasi, K., Yoneyama, M., Nishihori, T., Hirai, R., Kumeta, H., Narita, R., Gale, M., Jr., Inagaki, F. and Fujita, T. (2008). "Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses." **Mol Cell** 29(4): 428-440.
- Takaoka, A., Wang, Z., Choi, M. K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., Honda, K., Ohba, Y. and Taniguchi, T. (2007). "DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response." **Nature** 448(7152): 501-505.
- Takeuchi, O. and Akira, S. (2010). "Pattern recognition receptors and inflammation." **Cell** 140(6): 805-820.
- Tangye, S. G. and Tarlinton, D. M. (2009). "Memory B cells: effectors of long-lived immune responses." **Eur J Immunol** 39(8): 2065-2075.
- Taylor, G. (2003). "The phase problem." **Acta Crystallogr D Biol Crystallogr** 59(Pt 11): 1881-1890.
- Taylor, G. L. (2010). "Introduction to phasing." **Acta Crystallogr D Biol Crystallogr** 66(Pt 4): 325-338.
- Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2002). "Multiple sequence alignment using ClustalW and ClustalX." **Curr Protoc Bioinformatics** Chapter 2: Unit 2.3.
- Tschopp, J. and Schroder, K. (2010). "NLRP3 inflammasome activation: The convergence of multiple signalling pathways on ROS production?" **Nat Rev Immunol** 10(3): 210-215.
- 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." **J Immunol** 178(10): 6444-6455.
- Vilaysane, A. and Muruve, D. A. (2009). "The innate immune response to DNA." **Semin Immunol** 21(4): 208-214.
- Vitour, D. and Meurs, E. F. (2007). "Regulation of interferon production by RIG-I and LGP2: a lesson in self-control." **Sci STKE** 2007(384): pe20.
- Vonrhein, C., Blanc, E., Roversi, P. and Bricogne, G. (2007). "Automated structure solution with autoSHARP." **Methods Mol Biol** 364: 215-230.

- Walker, S. C., Avis, J. M. and Conn, G. L. (2003). "General plasmids for producing RNA in vitro transcripts with homogeneous ends." **Nucleic Acids Res** 31(15): e82.
- Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., Tuschl, T., Hartmann, G. and Patel, D. J. (2010). "Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I." **Nat Struct Mol Biol** 17(7): 781-787.
- Yanai, H., Ban, T., Wang, Z., Choi, M. K., Kawamura, T., Negishi, H., Nakasato, M., Lu, Y., Hangai, S., Koshiba, R., Savitsky, D., Ronfani, L., Akira, S., Bianchi, M. E., Honda, K., Tamura, T., Kodama, T. and Taniguchi, T. (2009). "HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses." **Nature** 462(7269): 99-103.
- Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y. M., Gale, M., Jr., Akira, S., Yonehara, S., Kato, A. and Fujita, T. (2005). "Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity." **J Immunol** 175(5): 2851-2858.
- 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." **Nat Immunol** 5(7): 730-737.
- Yu, J. W., Fernandes-Alnemri, T., Datta, P., Wu, J., Juliana, C., Solorzano, L., McCormick, M., Zhang, Z. and Alnemri, E. S. (2007). "Pyrin activates the ASC pyroptosome in response to engagement by autoinflammatory PSTPIP1 mutants." **Mol Cell** 28(2): 214-227.
- Yu, X., Acehan, D., Menetret, J. F., Booth, C. R., Ludtke, S. J., Riedl, S. J., Shi, Y., Wang, X. and Akey, C. W. (2005). "A structure of the human apoptosome at 12.8 Å resolution provides insights into this cell death platform." **Structure** 13(11): 1725-1735.
- 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(2): 315-330.
- Zhang, K., Kagan, D., DuBois, W., Robinson, R., Bliskovsky, V., Vass, W. C., Zhang, S. and Mock, B. A. (2009). "Mndal, a new interferon-inducible family member, is highly polymorphic, suppresses cell growth, and may modify plasmacytoma susceptibility." **Blood** 114(14): 2952-2960.
- Zhu, Z., Dumas, J. J., Lietzke, S. E. and Lambright, D. G. (2001). "A helical turn motif in Mss4 is a critical determinant of Rab binding and nucleotide release." **Biochemistry** 40(10): 3027-3036.

Abbreviations

List of commonly used abbreviations. Bases and amino acids (single or three letter code) as well as measures and units were abbreviated according to standard nomenclature.

5'PPP	5' Triphosphate
6-FAM	6-Carboxyfluorescein
aa	amino acids
AF488	AlexaFluor488
AIM2	Absent in melanoma 2
as	anti sense strand
AP-1	Activator protein 1
ASC	Apoptosis-associated Speck-like protein containing a CARD
Bcl-10	B-Cell Lymphoma 1 β
BIR	Baculoviral inhibitor of apoptosis proteins repeat
BSA	Bovine serum albumine
CARD	Caspase activation and recruitment domain
CARDIF	CARD adaptor inducing interferon- β
C/EBP β	CCAAT enhancer binding protein β
CLR	C-type lectin receptor
COP	CARD only protein
DAI	DNA-dependent activator of IRF
DAMP	Danger/damage-associated molecular pattern
DDX3	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3
DF	Death-fold
dsRNA/DNA	doublestrand RNA/DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMCV	Encephalomyocarditis virus
EMSA	Electrophoretic mobility shift assay
FADD	Fas-associated via death domain
FF	Fast flow resin
FINDII	F-interacting domain/domain with a function to find
FL	Full length
fwd	forward primer
GST	Glutathion-S-Transferase
HCV	Hepatitis C virus
HDV	Human delta virus
HEPES	N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
HFIP	Hexafluoroisopropanol
HIN200	Hematopoietic interferon-inducible nuclear proteins with a 200-amino acid repeat
HMGB	High mobility group box
hp	hairpin
HTS	High throughput screen
IEC	Ion exchange chromatography
iE-DAP	γ -D-glutamyl-meso-diaminopimelic acid
IFN	Interferon
I κ B	Inhibitor of NF- κ B
IKK β	I kappa B kinase β
IL	Interleukin
IPAF	Interleukin-1 β converting enzyme Protease Activating Factor
IPS-1	Interferon-beta promoter stimulator protein 1
IRF	Interferon regulatory factor
IVT	In vitro transcription
LGP2	Laboratory of genetics and physiology 2
LPS	Lipopolysaccharide
LRR	Leucine-rich repeat
MaR	Mannose receptor
MAVS	Mitochondrial anti-viral signaling protein
MDA5	Melanoma differentiation-associated gene 5
MDP	Muramyl dipeptide
MES	4-Morpholineethanesulfonic acid
MINCLE	Macrophage-inducible C-type lectin

Abbreviations

MPD	Methylpentanediol
MR	Molecular replacement
MsrB	Methionine sulfoxide reductase
MSS4	Guanine nucleotide exchange factor MSS4
MW	Molecular weight
NACHT	Domain found in NAIP, CIITA, HET-E and TP1
NAIP	Neuronal apoptosis inhibitory protein
NALP	NACHT domain- leucine-rich repeat-, and PYD-containing protein
NAP1	NF- κ B-activating kinase-associated protein 1
NBD	Nucleotide binding domain
NBS	Nucleotide binding site
NEMO	NF-kappa-B essential modulator
NF- κ B	nuclear factor "kappa-light-chain-enhancer" of activated B-cells
NLR	Nod-like receptor
NLRB	NOD-like receptor family, BIR domain containing
NLRC	NOD-like receptor family, CARD domain containing
NLRP	NOD-like receptor family, pyrin domain containing
NLS	Nuclear localization sequence
NOD	Nucleotide binding and oligomerization domain
nt	nucleotides
NTA	Nitrilotriacetic acid
OB	Oligonucleotide/-saccharide binding domain
PAGE	Polyacrylamide electrophoresis
PAMP	Pathogen-associated molecular pattern
PEG	Polyethyleneglycol
pI	Isoelectric point
Pol III	RNA Polymerase III
POP	PYD only protein
pRb	Retinoblastoma protein
PRR	Pattern recognition receptor
PYCARD	PYD and CARD domain-containing protein
PYD	Pyrin domain
RD	Regulatory (repressor) domain
rev	reverse primer
RIG-I	Retinoic acid inducible gene I
RIP	Receptor interacting protein/serine-threonine kinase 1
RLH	RIG-I like helicase
RLR	RIG-I like receptor
RMS	Root mean square
ROS	Reactive oxygen species
RVL	Rabies virus leader
s	sense strand
SAD	Single-wavelength anomalous dispersion
SAP130	Histone deacetylase complex subunit - Sin3-associated polypeptide p130
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography - gel filtration
SINTBAD	Similar to NAP1 TBK1 adaptor
SIRAS	Single isomorphous replacement with anomalous scattering
TAE	Tris acetic acid EDTA
TANK	TRAF family member-associated NF- κ B activator
TB	Tris borate
TBK	TANK-binding kinase 1
TCEP	Tris(2-carboxyethyl)phosphine
TEA	Triethylamin
TGS	TRIS Glycine SDS
TLR	Toll-like receptor
TRADD	Tumor necrosis factor receptor type 1-associated DEATH domain protein
TRAF	TNF Receptor associated factor
TRIM	Tripartite motif-containing
TRIS	Tris(hydroxymethyl)-aminomethan
VISA	Virus-induced signaling adaptor
wt	wildtype

Curriculum Vitae

Name:		Diana Angela Pippig
Date of birth:		11.02.1983
Place of birth:		Plauen, Germany
Nationality:		German
Education:	06/2001	Abitur at the Diesterweg-Gymnasium in Plauen
Academic training:	02/2007 – present	PhD candidate in Biochemistry at the Ludwig - Maximilians - Univeristy of München in the structural biology laboratory of Prof. Dr. Karl-Peter Hopnfer
	01/2006 – 12/2006	Research assistant in the protein crystallography laboratory of Dr. Sigurd Wilbanks at the University of Otago, New Zealand
	10/2001 – 12/2005	Studies of Biochemistry at the University of Bayreuth

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work on these fascinating and challenging projects in the still emerging field of innate immunity. I am particularly grateful for him being approachable at all times and for constant inspiring, constructive discussion and advice. I very much appreciated to conduct my PhD research in the stimulating environment of his lab.

I would further like to thank all members of the Hopfner laboratory with all my heart for keeping up such a great and creative atmosphere in the lab, which makes working there truly enjoyable and the occasional lab frustration so much more bearable. Especially I would like to acknowledge Manu, Alex, Brigitte and Olga for all the small and big things they do in the daily lab routine and beyond. I would also like to thank everybody who contributed to my projects. Special thanks go to Katja, Sophia and Gregor for great scientific discussions and all sorts of help.

I am grateful for having been given the opportunity to be a member of the International Max Planck Research School for Molecular and Cellular Life Sciences and the great scientific exchange and training this provided. In this context I would also like to thank Prof. Dr. Elena Conti and Prof. Dr. Karl-Klaus Conzelmann for joining my TAC, constructive discussions and great counsel.

Ich möchte mich außerdem bei Christoph, meinen Freunden, meiner Familie und ganz besonders und von ganzem Herzen bei meinen Eltern bedanken die während meiner Doktorarbeit ein starker Rückhalt waren und mich immer und überall unterstützen.