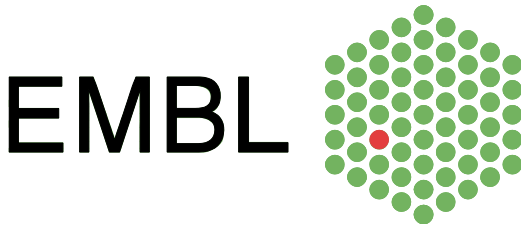
Structural studies of host-virus interactions looking at two examples: the innate immunity receptor RIG-I and the influenza virus RNA polymerase endonuclease

Eva Kowalinski



Structural studies of host-virus interactions looking at two examples: the innate immunity receptor RIG-I and the influenza virus RNA polymerase endonuclease

Doctoral Thesis of Eva Kowalinski

Joint PhD of the European Molecular Biology Laboratory and the University Joseph Fourier Grenoble

Structural Biology and Nanobiology

Public defense 9 December 2010

Jury:	**Winfried Weissenhorn**	**Jury President**
	Daniel Kolakofsky	**Rapporteur**
	Denis Gerlier	**Rapporteur**
	Stephen Cusack	**Thesis Director**

Études structurales des interactions virus-hôte à travers deux exemples : le récepteur du système immunitaire inné RIG-I et le domaine endonucléase de l'ARN polymérase du virus de la grippe

Thèse de doctorat présentée par Eva Kowalinski

Doctorat conjoint entre le Laboratoire Européen de Biologie Moléculaire et l'Université Joseph Fourier Grenoble

Biologie structurale et nanobiologie

Soutenance publique le 9 Décembre 2010

Jury:	**Winfried Weissenhorn**	**Président**
	Daniel Kolakofsky	**Rapporteur**
	Denis Gerlier	**Rapporteur**
	Stephen Cusack	**Directeur de Thèse**

If we knew what it was we were doing,
it would not be called research, would it?
Albert Einstein

Summary

The first line of defense against invading pathogens in the human body is the innate immune system. Astonishingly, with only a handful of different pathogen recognition receptors (around 50), the innate immune system is able to detect a remarkably broad variety of pathogen specific molecules to trigger protective pathways and to activate the adaptive immune system. In the case of intruding viruses, two families of pattern recognition receptors (PRRs) are active: retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs) and Toll-like receptors (TLRs). The family of RIG-I like receptors includes the proteins RIG-I, MDA5 and LGP2, which recognize viral RNA in the cytosol. In the first part of this thesis, aspects of the RIG-I pathway are discussed: With which RNA does RIG-I interact and how? Does the oligomeric state of RIG-I change upon RNA binding in order to trigger signaling? How is RIG-I regulated by ubiquitin and its E3 ubiquitin ligase TRIM25? The structure of the PRYSPRY domain of TRIM25, its putative RIG-I binding domain, will be presented. Furthermore, preliminary work on the second receptor MDA5 in complex with parainfluenza virus V, which inhibits the MDA5 pathway, protein will be shown.

One of the activators of the receptor RIG-I is the RNA of influenza virus. Influenza viruses belong to the family of *Orthomyxoviridae* that affect birds and mammals and spread in seasonal epidemics. In pandemic years, this can result in up to millions of deaths worldwide, underlining the need for research on efficient novel anti-viral drugs. In the second part of the thesis (appended as an article manuscript), the A/California/04/2009-H1N1 ”swine flu” influenza RNA polymerase will be investigated as a novel antiviral drug target. Crystal structures of the endonuclease domain (PA-Nter) of the polymerase with four different inhibitors are presented. Moreover, the atomic structures of H1N1 PA-Nter with rUMP and dTMP, elements of the nucleic acid substrate, in the active site are discussed. These high resolution structures will serve as a basis for structure based inhibitor optimization.

Résumé

Le système immunitaire inné constitué la première barrière de défense du corps humain contre les agents infectieux. Constitue d'une poignée de récepteurs seulement (approximativement 50), ce système est capable de détecter la plupart des agents pathogènes et d'activer le système immunitaire adaptatif. Dans le cas d'une infection virale, deux classes de récepteurs sont mobilisées : les RLRs (Retinoic acid Like Récepteurs) et les TLRs (Toll Like Receptors). La famille des RLRs comprend notamment trois protéines, RIG-I, MDA-5 et LGP2, qui reconnaissent la présence d'ARN viral dans le cytosol. Dans la première partie de cette thèse, nous discuterons de la voie d'activation de RIG-I : avec quels type d'ARN et de quelle manière RIG-I interagit-il ? L'état d'oligomerisation de RIG-I change-t-il lors de l'interaction avec l'ARN ? De quelle manière RIG-I est-il influencé par l'ubiquitine et sa ligase E3, TRIM25 ? Nous présenterons également la structure du domaine PRYSPRY de TRIM25, domaine minimum nécessaire à l'interaction avec RIG-I. Nous décrirons également les travaux préliminaires obtenus sur un complexe constitué de MDA5 et une molécule d'inhibition virale.

L'ARN des virus de la grippe est un des activateurs du récepteur RIG-I. Les virus de la grippe appartiennent à la famille des *Orthomyxoviridae* qui touche les mammifères et les aves créant des épidémies saisonnières. Ces dernières peuvent causer jusqu'à plusieurs millions de morts lors de pandémies, soulignant le besoin de trouver de nouvelles solutions thérapeutiques. Dans la seconde partie de ma thèse, nous nous intéresserons à l'activité endonuclease de l'ARN polymerase du virus A/California/04/2009-H1N1 de la grippe porcine comme une cible pour de nouvelles molécules antivirales. Les structures de quatre inhibiteurs en complexe avec ce domaine seront présentés. Nous présenterons également la structure de PA-Nter avec les substrats rUMP et dTMP co-cristallisés dans le site actif. Toutes ces structures atomiques forment une base pour l'optimisation et la synthèse d'inhibiteurs.

Zusammenfassung

Die erste Verteidigungsline gegen Krankheiserreger im menschlichen Körper ist das Angeborene Immunsystem. Erstaunlicherweise genügt eine geringe Anzahl von Rezeptoren (ungefähr 50), um Erreger verschiedenster Art aufzuspühren, schnelle Schutzmechanismen in die Wege zu leiten und letztendlich die Erworbene Immunität zu aktivieren. Im Falle einer Virusinfektion sind zwei Familien von "Pattern-Recognition" Rezeptoren (PRRs) aktiv: RIG-I ähnliche Rezeptoren (RLRs) und Toll-ähnliche rezeptoren (TLRs). Die Familie der RIG-I ähnlichen Rezeptoren umfasst drei Proteine, RIG-I, MDA-5 and LGP2, die gemainsam virale ribonuleinsäuren (RNA) im Zytosol aufspühren. Der erste Teil dieser Doktorarbeit setzt sich mit dem RIG-I Signalweg auseinander: An welche RNA bindet RIG-I? Oligomerisiert der Rezepor wenn er RNA erkannt hat? Und welche Rolle spielt die Interaktion zwischen RIG-I und Ubiquitin, beziehungsweise RIG-I und seiner E3 Ubiquitin ligase TRIM25? Die Kristallstruktur der PRYSPRY Domäne von TRIM25, die vermeintliche Bindungsdomäne zu RIG-I, wird vorgestellt. Desweiteren werden vorläufige Ergebnisse von Arbeiten am zweiten RLR Rezeptor, MDA5, im Komplex mit einem viralen Hemmer, Parainfluenza Virus Protein V vorgestellt.

Unter den viralen RNAs, die RIG-I erkennt, befindet sich auch die RNA von Influenzaviren ("Grippeviren"). Influenza gehört zu den *Orthomyxoviridae*, die sowohl Säugetiere als auch Vögel infizieren und breitet sich in Grippeepidemieen aus, was in Pandemiejahren zu Millionen von Toten führen kann. Pandemien, wie die "Schweinegrippe" 2009, führen die Notwendigkeit für Forschung an effizienten anti-viralen Medikamenten vor Augen. Im zweiten Teil dieser Doktorarbeit (als Publikationsmanuskript angegliedert) wird die Endonuclease der RNA Polymerase (PA) des Schweinegrippe-Erregerstamms A/California/04/2009-H1N1 als Angriffsziel für neue anti-virale Medikamente untersucht. Kristallstrukturen der Endonuklease-Domäne mit vier verschiedenen anti-viralen Molekülen und rUMP und dTMP im aktiven Zentrum wurden gelöst und können nun zur Optimierung der Molekülstrukturen engesetzt werden.

Contents

Chapter 1

Introduction

Summary of introduction

Viruses are small infectious agents replicating only inside living cells. They infect all types of living organisms. However, cells have developed different mechanisms to protect agianst viral infection. Mammals posses an immune system that constantly provides protection from invading pathogens. One of the anti-viral signaling cascades in the innate immune systems is triggered by a family of three receptors that detect viral RNA in the cytosol: RIG-I, MDA5 and LGP2. RIG-I and MDA5 detect different sets of viruses. LGP2 seems to have a regulatory function in the signaling. The receptor RIG-I is the best studied of the three and binds specifically to short blunt-end double strand 5'-tri-phosphorylated RNAs (among those the RNA of influenza viruses). RIG-I signaling is further regulated by poly-ubiquitin chains synthesized by the E3 ubiquitin ligase TRIM25. Additional regulation of the pathway is achieved by other ubiquitinases and specific phoshorylation of RIG-I.

Influenza viruses belong to the family of *Orthomyxoviridae* that affect birds and mammals. Influenza spreads in seasonal epidemics. The recent pandemic influenza strain of 2009, the H1N1 "swine flu", combines genes from human, pig, and bird flu. The most effective treatment of influenza is the prevention by vaccination, but the rapidly evolving virus requires constant adaptation of vaccines and makes previsions difficult. To date, anti-influenza drugs available on the market include neuraminidase inhibitors and M2 inhibitors. The endonuclease subunit of the RNA polymerase of influenza virus is responsible for cleaving off capped primers from cellular mRNAs in a mechanism called "cap-snatching". As cap-snatching is essential for translation of viral transcripts, this process is a potential target of novel anti-viral drugs.

Résumé du chapître “Introduction”

Les virus sont des agents infectieux pouvant se répliquer à l'intérieur des cellules vivantes et possèdant la faculté d'infecter différents types d'organismes. Cependant, certaines cellules ont développé différents mécanismes de protection afin de prévenir l'infection virale. Les mammifères disposent ainsi d'un système immunitaire leur procurant une protection constante et efficace contre de nombreux pathogènes. Une des voies de signalisation antivirale du système immunitaire inné est déclenchée par l'activation d'une famille de trois récepteurs reconnaissant la présence d'ARN viral au niveau du cytosol : RIG-I, MDA5 et LGP2. Une fois activés, ils déclenchent une cascade de signalisation conduisant à la production d'interférons. RIG-I et MDA5 sont sensibles à différents types de virus. LGP2, quant à lui, semble plus avoir un rôle inhibiteur sur la transmission du signal. Le récepteur RIG-I, le plus étudié des trois, est capable d'interagir spécifiquement avec les courtes séquences d'ARN double brin tri-phosphorylées en 5' (parmi tous les ARN du virus de la grippe). La voie RIG-I est régulée par l'ubiquitine ligase TRIM25 qui est directement responsable de son ubiquitination mais aussi synthétise des chaines poly-ubuitinées qui interagissent avec RIG-I. Des voies de contrôle supplémentaire de RIG-I impliquent plusieurs ubiquitinases et des kinases spécifiques. Les virus de la grippe appartiennent à la famille des *Orthomyxoviridae* qui touche les mammifères et les oiseaux. Le virus de la grippe se propage lors d'épidémies saisonnières. L'épidémie de grippe aviaire de 2009 (A-H1N1), est le fruit d'une combinaison de gènes humain, porcin et aviaire. A ce jour, le traitement préventif le plus efficace est la vaccination mais la rapide évolution du virus nécessite une adaptation constante des vaccins et rend les prédictions d'évolution difficiles. A ce jour, certaines molécules anti-virales sont disponibles sur le marché comme des inhibiteurs de la neuraminidases ou encore des inhibiteurs de la protéine M2. La sous-unité portant l'activité endonucléase de l'ARN polymérase de virus de la grippe est responsable de la coupure de la coiffe des ARNm cellulaires, ce mécanisme est appelé ”cap-snatching ”. Ce dispositif, essentiel à la transcription des ARN viraux, est une cible potentielle pour des nouvelles molécules antivirales.

1.1 Strategies of the innate immune system

The innate immune system is the first line of defense that provides immediate protection against intruding pathogens. Innate immunity is found in all classes of plant and animal life. Cells of the innate immune system recognize and respond to a broad spectrum of pathogen antigens in an unspecific generic way. The innate immune system operates via various mechanisms, i. e. (i) phagocytotic cells (macrophages, neutrophiles, dendritic cells) attracted to the source of inflammation, (ii) the complement system, which consists of small proteins circulating in the blood, and (iii) the interferon system which is a warning system to viral infections. Finally, the innate immune system invokes the adaptive immunity by antigen presentation to assure long lasting protective immunity.

1.1.1 Pattern recognition

The crucial mechanism of the innate immunity is to protect the organism against all kinds of known and unknown microbes. Detection of pathogen-associated molecular patterns (PAMPs), conserved molecules associated with groups of pathogens, by the immune system is mediated by pathogen recognition receptors (PRRs). A schematic overview of pattern recognition receptors in the cell is given in figure 1.1. PAMPs include bacterial carbohydrates (e.g. lipopolysaccharide or LPS, mannose), nucleic acids (e.g. bacterial or viral DNA or RNA), bacterial peptides (flagellin), peptidoglycans and lipotechoic acids (from gram positive bacteria), N-formylmethionine, lipoproteins and fungal glucans. These general antigens are recognized by a variety of PRRs like the membrane bound Toll-like receptors (TLRs) and the mannose receptor [12]. Furthermore there are extracellular PRRs secreted to the extracellular space and cytosolic PRRs like NOD-like receptors (NLRs) and RIG-I like receptors (RLRs). For a review on pattern recognition receptors see Takeuchi and Akira, 2010 [140].

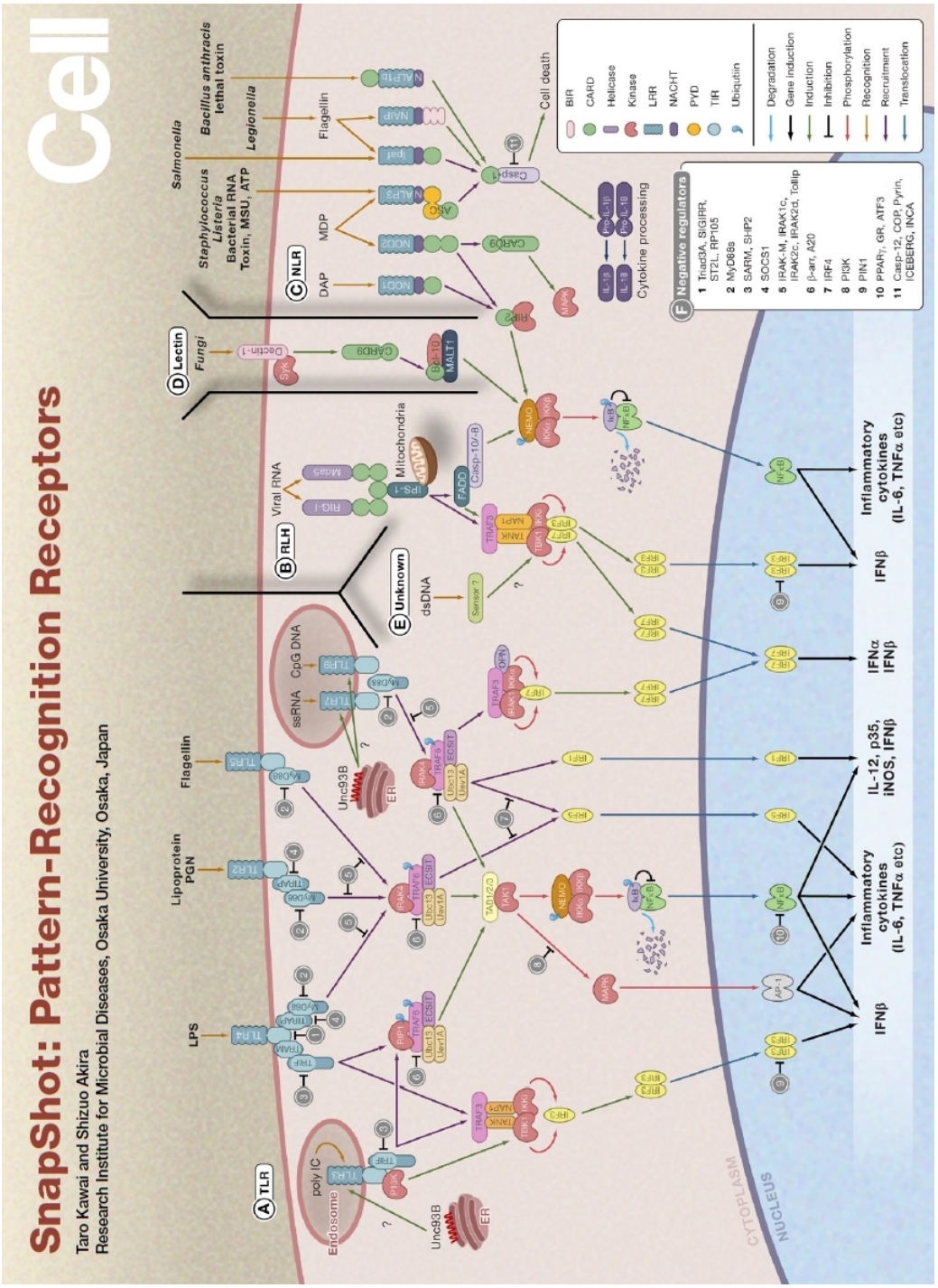


Figure 1.1: *A general overview on pattern recognition receptors [68].*

1.1.2 Defense against viruses

Viral infections provoke an inflammatory response of the mammalian cell that in turn releases cytokines and interferons (IFNs) into the extracellular space. The cytokines then activate immune cells, up-regulate antigen presentation in infected and tumor cells and increase the ability of yet uninfected host cells to resist the infection. Typically, viruses are detected by PRRs that recognize their characteristic nucleic acids. Recognition receptors for microbic RNA and DNA include TLRs (TLR3, TLR7, TLR8 and TLR9) bound to the endosomal membrane of phagocytotic cells (with the exception of TLR3 which is found also in non-immune cells), the cytosolic nucleic acid receptors AIM2 (absent in melanoma 2) [58], DAI (DNA-dependent activator of IRF-regulatory factors) [137] and the family of RIG-I like receptors (RLRs). In particular, PRRs for viral RNA include (i) TLR3 recognizing double stranded viral RNA and polyI:C (a synthetic ligand mimicking viral RNA), (ii) TLR7 sensing single stranded viral RNA and (iii) RLRs that are cytosolic receptors for intracellular viral RNA [67]. Unlike TLRs which are only present in cells of the immune system, RLRs are present in all type of cells in the organism. A schematic overview of receptors for foreign nucleic acids is given in figure 1.2.

In plants and invertebrates, RNA interference serves as an antiviral defense mechanism. Here, double microbial stranded RNA is detected by enzymes of the Dicer family and cleaved into small interfering RNAs (siRNA) of 21-23 nucleotides inhibiting viral protein expression on the mRNA level.

1.2 RIG-I like receptors

1.2.1 The family of RIG-I like receptors are receptors for viral RNA

RIG-I like receptors (RLRs) are cytosolic recognition receptors for viral RNA. The three members of this protein family are: retinoic acid inducible gene I (RIG-I), encoded by the gene Ddx58, melanoma differentiation associated gene 5 (MDA5), encoded by the

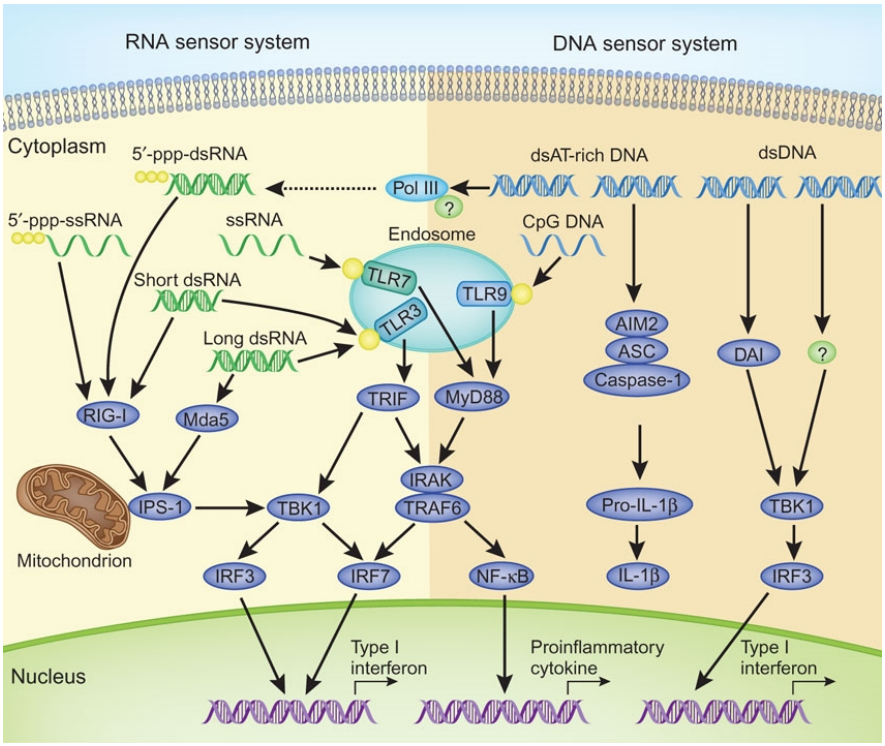


Figure 1.2: *Recognition mechanisms of foreign nucleic acids in the innate immune system [16].*

gene Ifih1, and laboratory of genetics and physiology 2 (LGP2), encoded by the gene Dhx58. RIG-I and MDA5 sense intracellular viral RNA and induce a signaling cascade leading to an interferon response of the cell. RIG-I is a 106 kDa protein consisting of two C-terminal caspase recruitment domains (CARDs) followed by a DExD/H-box RNA helicase domain and a C-terminal domain (CTD), sometimes referred to as "repressor" domain (RD) [120]. The CTD contains a zinc binding site [146, 83, 78]. RIG-I helicase domain is an ATPase and has an ATP powered dsRNA translocation activity preferentially on dsRNA [92]. The order of domains in the second receptor, MDA5 (117 kDa), is identical; it shares 23% sequence identity with the CARDs and 35% identity with the helicase domain of RIG-I. The two receptors have a different range of viruses to which they respond (see section 1.2.4). The receptor LGP2 (77 kDa) is comprised of a helicase domain and a CTD, but is lacking the CARDs. It shares 31% and 41% identity with the RNA helicase domains of RIG-I and MDA5, respectively. LGP2 has a regulatory role in the anti viral responses [143, 145, 122].

1.2.2 Domain architecture of RIG-I like receptors

The caspase activating and recruitment domain (CARD)

The CARD motif belongs to the family of death fold protein motifs. This fold comprises six to seven anti-parallel α-helices with a hydrophobic core and charged residues on the surface in a Greek key topology (figure 1.4 A). Other members of this family are pyrin domain (PYD), death domain (DD) and death effector domain (DED). These domains function primarily in the regulation of apoptosis and inflammatory responses by mediating protein-protein interactions via homeotypic contacts (CARD-CARD, DED-DED etc.). Those complexes are often highly multimeric and signaling is achieved by receptor clustering [85, 13], like for example in the crystal structures of the MyD88-IRAK4-IRAK2 complex and the Fas/FADD complex [80, 119, 127]. Several structures of single CARDs have been solved, but little is known about the exact stoichiometry of the various CARD-CARD complexes. To date, there are CARD complex structures available of Apaf-1 with caspase-9 and their *C. elegans* homologues CED-4-CED-9 [112, 156, 153]. However, NOD1 CARD, which is a monomer in solution [84] (figure 1.4 A), crystallizes as a dimer [26, 132]. Further studies revealed that the conformational arrangement in helix α6 which enables dimerization is pH dependent and an increasing acidity favors dimer formation. The folding kinetics of CARD domains are complex, as studies on Apaf-1 CARD, RICK-CARD and CP1-CARD show [19, 20, 87]. This might be a reason for difficulties to express and purify recombinant CARDs. The CARDs of RIG-I have an inhibitory effect on the ATPase function, as mutants lacking the CARDs have an elevated ATPase activity and translocate faster along RNA strands [92, 28]. The structure of the N-terminal single CARD domain of the mitochondrial adapter of RIG-I, MAVS, has been solved by X-ray crystallography [110]. For a review on death fold motives in immune signaling see Monie *et al.* 2009 [99].

The RNA helicase domain

Helicases are a class of motor proteins that move directionally along a nucleic acid phosphodiester backbone, separating two annealed nucleic acid strands (i.e. DNA, RNA, or RNA-DNA hybrids) under consumption of energy from ATP hydrolysis. As often misunderstood, helicases do not necessarily have an unwinding activity for DNA or RNA double strands [63]. RIG-I like receptors (RLRs) belong to the family of DExH/D box RNA helicases sharing nine sequence motives: Q-motif, motif 1, motif 1a, motif 1b, motif II, motif III, motif IV, motif V, and motif VI. Motif II is also known at the Walker B site and contains the amino acids DExH/D (DEAD in the case of RLRs), which gives its name to this helicase family. Motif 1, motif II, the Q motif, and motif VI are needed for ATP binding and hydrolysis, while motifs 1a, 1b, III, IV, and V may be involved in intramolecular rearrangements and RNA interaction. An overview on the DEAD helicase motives and their functions is given in figure 1.3. K270 is the crucial residue for ATP hydrolysis in RIG-I. K270A and D372 mutants are inactive in antiviral signaling, even though they retain RNA binding ability [154, 108, 157]. To date, the function of the CARDs in signaling and the CTD in 5'-PPP binding seems clear, but the meaning of ATPase- and translocase activity of the helicase of RIG-I remains to be understood.

A crystal structure of *Pyrococcus furiosus* Hef helicase domain, a close homologue of the helicase domain of RIG-I is shown in figure 1.4 [93]. The helicase domain consists of three structurally distinct units. Two units mark the conserved moiety which comprises the typical helicase sequence motifs mentioned above. The two helicase parts, in the sequence separated by an insertion but structurally located in close proximity, have an α/β fold with a parallel β-sheet on which α-helices pack from both sides. The third "inserted" unit consists of α-helices and varies between different RNA helicases. The units of the helicase are flexibly connected by flexible linkers. Further members of the DEAD box family are for example the translation initiation factor eIF4A and the endoribonuclease Dicer [62, 25, 81, 56].

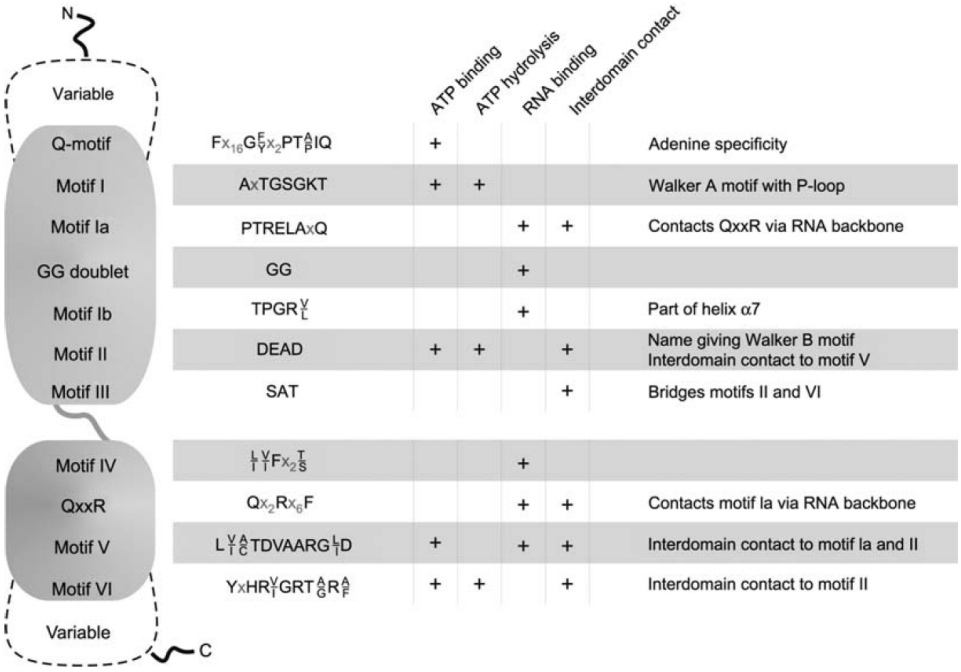


Figure 1.3: *The motives which characterize DEAD box helicases [56].*

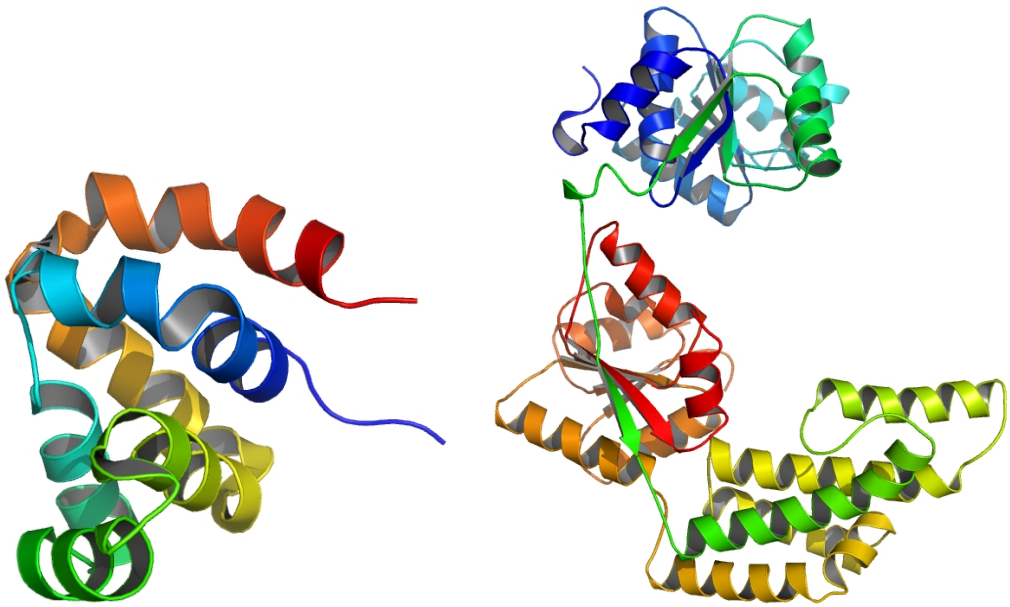


Figure 1.4: *(**A**) Solution structure of the NOD1 caspase activating and recruitment domain (CARD) [84]. (**B**) Crystal structure of Pyrococcus furiosus Hef helicase domain, which shares 22% homology with the helicase domain of RIG-I [93]. Both structures are colored from N to C-terminus in the order: blue, green, yellow, orange, read*

The C-terminal domain of RIG-I like receptors

The C-terminal domain (CTD) is responsible for RNA binding and is present in all three receptors, RIG-I, MDA5 and LGP2 [28, 138, 83]. Expression of CTD alone inhibited IFN signaling *in vivo* while a RIG-I mutant lacking the CTD constitutively activates IFN promoters [120]. Crystallographic and NMR high resolution structures are available for all of them (PDB coordinates and references are listed in section 1.2.7). The domain consists of a twisted anti-parallel main β-sheet and a smaller second anti-parallel β-sheet. The sheets are stabilized by surrounding loop regions and short helices. The domain forms a globular structure. One Zn^{2+} ion is coordinated close to the smaller β sheet. The RNA binding site has been localized to a large positively charged patch on the surface of the main β sheet of all three molecules (figure 1.5) [138].

RIG-I and LGP2 CTDs were both co-crystallized in a complex with dsRNA (Figure 1.6). In both structures, one protein molecule is bound to each 5' end of the dsRNA to form a 2:1 protein:dsRNA complex [146, 83, 78]. In RIG-I, the positively charged residues of the RNA binding patch interact with the 5'-PPP through electrostatic interactions, but also the rest of the RNA interacts extensively with other parts of the CTD by all kinds of interactions (hydrophobic, electrostatic, hydrogen bonds and aromatic base stacking). However, the complex structures of RIG-I bound to 5'-PPP dsRNA and LGP2 bound to blunt-end dsRNA show significant differences. The interaction surface of LGP2 CTD with RNA is a about 30% larger than in the case of RIG-I CTD. Since RIG-I has also a nanomolar affinity to blunt-end dsRNA without 5'-PPP, it is speculated that the mode of binding in that case might resemble the one in the LGP2 CTD complex structure. It would be interesting to see the vice-versa complex structures. In the crystal structure, RIG-I CTD interacts with the first four nucleotides at the 5'end of the dsRNA, with only limited contribution of the complementary strand. Residues K888, K858, K861 and H847 are critical for 5'-PPP recognition; K907 and K909 interact with the phosphate backbone of the dsRNA and further residues (F853, H830, Y831, I887, S906, W908) interact with

the RNA in an sequence unspecific manner [83, 146, 28, 139]. Despite of K888, the residues responsible for 5'-PPP binding are not conserved in MDA5 and LGP2, which makes RIG-I the only receptor specific for the triphosphate RNA. The CTD of MDA5 differs from the ones of RIG-I and LGP2 since the predicted RNA binding patch is flatter and more open. MDA5 does not need the 5'-PPP feature to recognize viral RNA. A complex structure of MDA5 CTD with RNA is not yet available and the exact ligand for MDA5 remains to be determined.

1.2.3 The signaling cascade of RIG-I

Inactive RIG-I has a closed conformation, in which the CARDs are inaccessible for signal transduction [120, 140, 155, 157]. Upon activation of RIG-I it is supposed to switch from a closed to an open conformation and to oligomerize [141, 154, 141, 154, 28, 77, 125, 120, 113]. For full activation of RIG-I, two triggers are needed: (i) a viral RNA ligand and (ii) an ubiquitination signal [44, 157]. Zeng and co-workers propose a multistep mechanism for the complete activation of RIG-I (figure 1.7) [157]. First RIG-I binds to the viral 5'-PPP with its CTD [43]. In a second step the RNA binding causes dimerization of RIG-I and activates ATPase activity [28]. The ATPase needs magnesium as a co-factor. ATPase function enables the translocation of RIG-I along the RNA [92]. The translocating RIG-I dimer is supposed to have an open conformation with accessible CARDs that form a cluster for binding K63-poly-ubiquitin chains. ATPase activity of RIG-I is necessary for interaction with ubiquitin. Poly-ubiquitin binding may cause additional conformational changes or oligomerization that enables binding to the adapter molecule MAVS in a final step [157]. However, Gack *et al.* found direct ubiquitination of the CARDs of RIG-I by the ubiquitin ligase TRIM25 necessary for MAVS activation [44, 46]. MAVS (mitochondrial antiviral signaling) (also referred to as CARDIF/IPS-1/-VISA) [154, 129, 69, 152, 86, 136] and RIG-I interact through their CARDs. MAVS occurs either bound to the outer mitochondrial membrane via a C-terminal transmem-

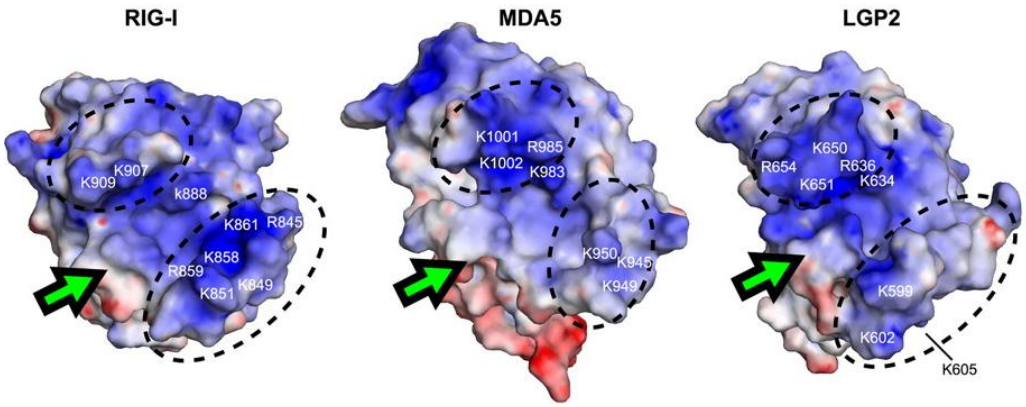


Figure 1.5: *The positive charged RNA binding patch on the CTD of RIG-I, MDA5 and LGP2. Positive charges are colored in blue and negative charges in red on the surface representation of the domains [138].*

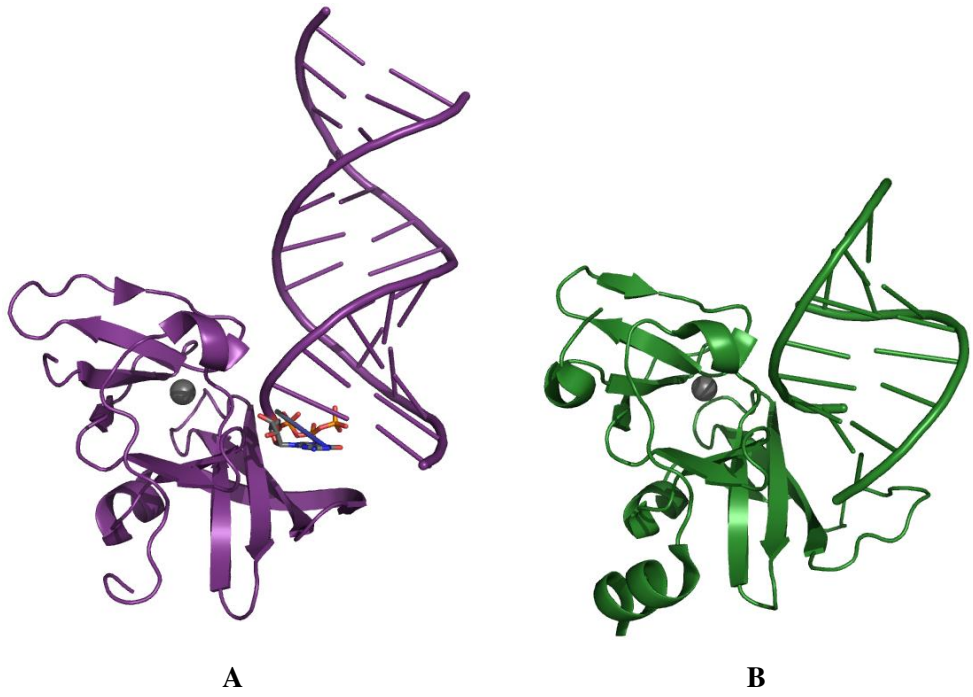


Figure 1.6: *CTDs bound to RNA. The zinc ion in each structure is marked in gray. (**A**) The CTD of RIG-I bound to a 12-mer of 5'-PPP dsRNA [83]. (**B**) The CTD of LGP2 bound to a 8-mer of blunt-end dsRNA [78].*

brane helix [129] or anchored to the peroxisomal membrane [36]. Upon activation by RIG-I, mitochondrial MAVS dimerizes [9] and induces an inflammatory response via two modes of action. (i) Firstly, via the adapters TRAF3, TANK and TRADD it relays the signal to the kinases TANK-binding kinase 1 (TBK1) and IκB kinase ε (IKKε) which subsequently activate the transcription factors interferon response factor 3 (IRF3), IRF7, and nuclear factor κB (NF-κB). These coordinate in the following the expression of type I IFN genes. (ii) Secondly, mitochondrial MAVS interacts via CARD-CARD interaction with the protein CARD9 to form a signaling complex with BCL10 to trigger the production of IL-6 and IL-1β via NF-κB. Mitochondrial MAVS is negatively regulated by NLRX1 [90], another mitochondrial membrane protein. Recently, MAVS has also been found on peroxisomal membranes. Peroxisomal MAVS triggers rapid interferon-independent expression of defense factor that provide short term protection while the mitochondrial MAVS responds in delayed but longer lasting kinetics [36].

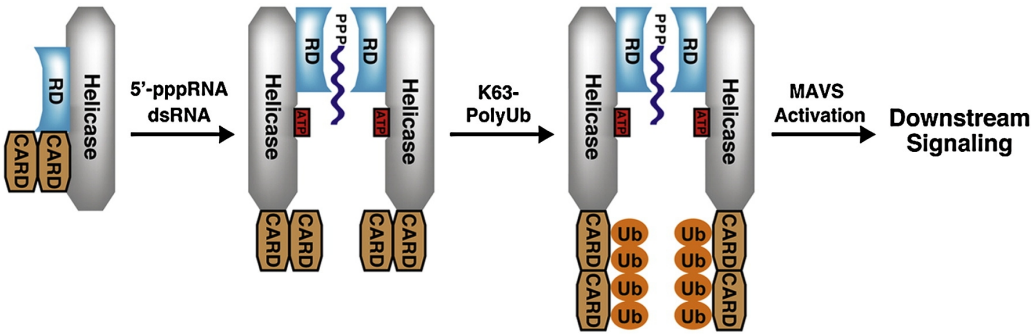


Figure 1.7: *Proposed activation mechanism of RIG-I by Zeng et al. [157]*

A further mode of action of RIG-I is described, in which it directly activates a caspase-1-activating multi protein complex, the ASC inflammasome. This process is independent of MAVS, CARD9 or TRIM25. The RIG-I ASC inflammasome then promotes the processing of proinflammatory cytokines like IL-1 by caspases [126]. Another adapter for RIG-I is the endoplasmatic reticulum bound protein STING [60]. In figure 1.8 a schematic overview of the different modes of action of RIG-I is drawn.

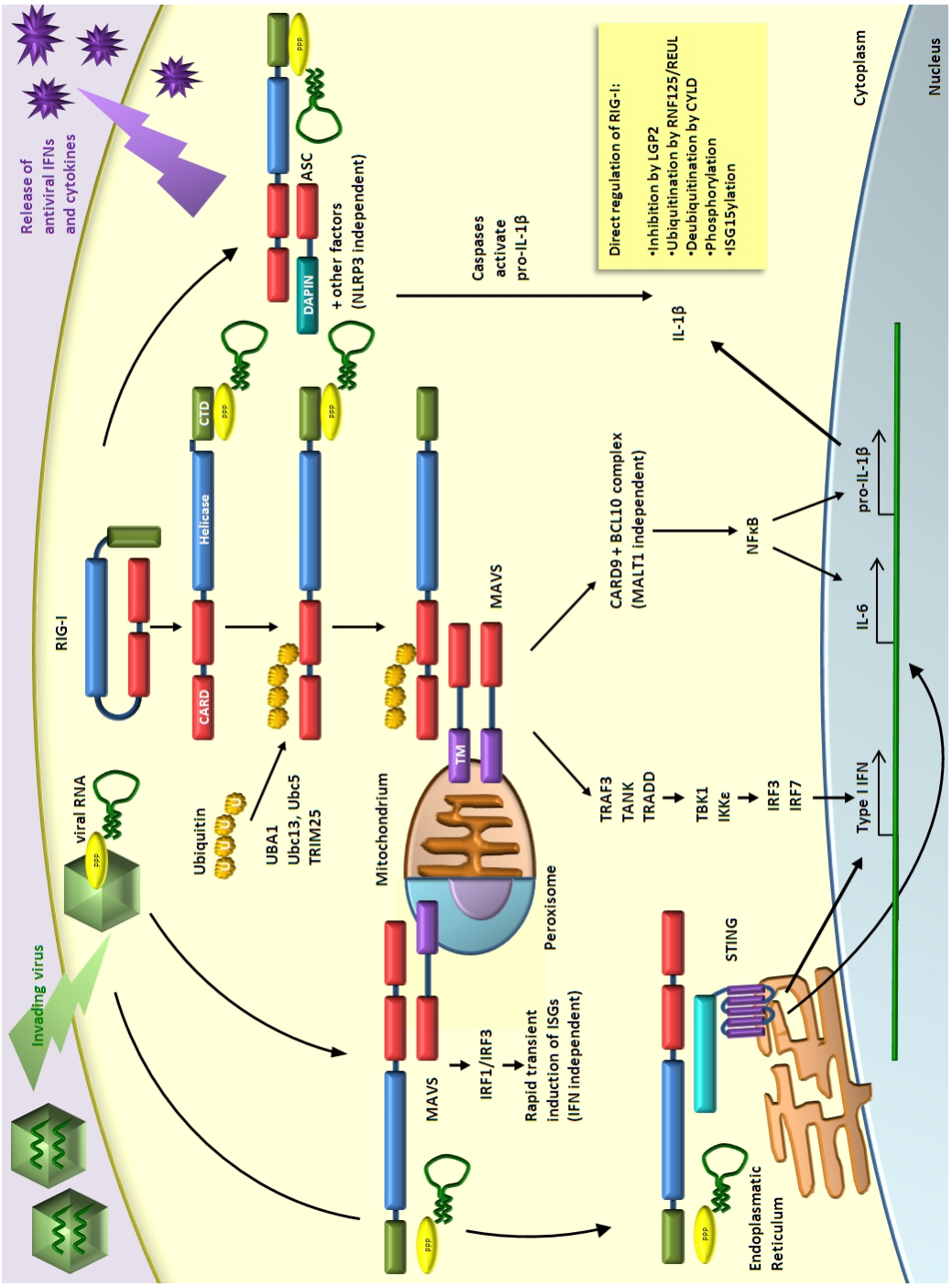


Figure 1.8: *Scheme of the RIG-I signaling pathways.*

1.2.4 RNA ligands for RIG-I

The nature of the RNA ligand required for RI-I signalling has been the subject of much investigation, controversy and confusion. While MDA5 seems critical for the detection of picornaviruses carrying RNAs with protected 5'ends (i. e. poliovirus, encephalomyocarditis or theiler's virus) [59, 105], RIG-I detects orthomyxovirus (such as influenza A virus) and paramyxovirus (such as measles, mumps), sendai virus and positive stranded viruses with unprotected 5'RNA genomes (like hepatitis C or Japanese encephalitis viruses). Mice deficient in RIG-I or MDA5 have impaired immune reactions to the corresponding set of viruses [66]. The ideal ligand of RIG-I seems to be a short blunt-ended 5'-PPP RNA with a double stranded portion at the 5'end of at least about 20 bases. Of those, at least the six base pairs adjacent to the 5'-end should perfectly match without bulges [124, 125, 123, 96]. Also poly I:C, a synthetic RNA double strand analogue and viral double strands can activate RIG-I [65, 55, 134]. Efforts to determine the relevant ligand *in vivo* in infected cells have been made, but contradictory studies found either the defective interfering particle (DI) or genomic viral RNA bound to RIG-I [11, 118].

There is evidence that RIG-I is additionally fed by RNA polymerase III transcripts [1, 23]. RNA polymerase III is involved in a DNA sensing pathways in the innate immune system and transcribes microbial DNA templates into dsRNA containing 5'-PPPs, which in turn activate RIG-I. This mechanism is independent of the DNA senors DAI (DNA-dependent activation of interferon regulatory factors) and AIM2 (absent in melanoma 2) [137, 58]. Those signaling molecules are themselves able to invoke a interferon response. The exact natural ligand for MDA5 has not yet been identified, but long double stranded RNAs and long stretches (longer than 200 bases) of polyionosinic-polycytidylic acid (polyI:C) can activate MDA5 mediated signaling [66, 49, 65].

1.2.5 RIG-I activity is fine tuned by post translational modifications

Further regulation of the pathway on the receptor level is achieved by ubiquitination and de-ubiquitination of RIG-I. Ubiquitin is a 8.5 kDa, highly-conserved regulatory protein. One ubiquitin molecule alone or a chain consisting of several ubiquitin units can be covalently linked to proteins as post-translational modification. There are three classes of enzymes involved in the modification of proteins by ubiquitin: an E1 ubiquitin activation enzyme which activates ubiquitin under consumption of ATP, an E2 ubiquitin-conjugating enzyme and finally an E3 ubiquitin ligase. The E3 ligase binds both E2 and the target substrate and covalently links the ubiquitin to the target protein in a substrate specific manner. The function of the ubiquitin modification depends on the number of ubiquitin units and the specific lysine residue by which the ubiquitins are linked. The ubiquitin signals label proteins for proteasomal degradation, control protein stability and function, or even influence intracellular localization of the protein (figure 1.9). In general, ubiquitin units linked via lysine 48 (K48) mark the protein for proteosomal degradation whereas K63-linked units play a role in cell signaling. K48 linked chains adopt a globular structure, while K63-linked chains have structural similarity to linear chains [72]. Poly-ubiquitin chains can also signal as non-covalently bound unanchored chains [157].

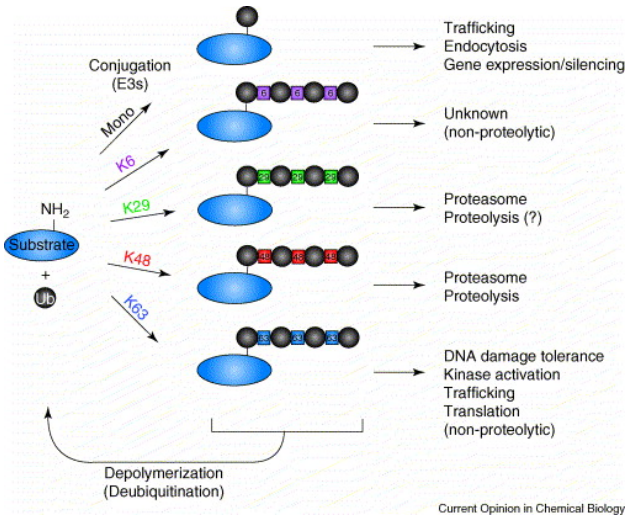


Figure 1.9: *Different roles of different ubiquitin chains [106].*

The E3 ubiquitinase TRIM25 (which will be described in detail in section 1.3) was described to influence RIG-I signaling [44, 46, 157], but the findings are partially contradictory. Gack and co-workers found that TRIM25 PRYSRPY domain directly binds to the first CARD of RIG-I (residue T55 of RIG-I is essential for binding) to ubiquitinate K172 which is located in the second CARD and that the ubiquitination of RIG-I is necessary for RIG-I signaling [44]. In contrast, Zeng *et al.* found that direct ubiquitination of RIG-I is not required. They propose that TRIM25 synthesizes, in concert with the E1 (UBA1) and E2 (Ubc13 and Ubc5) ligases, unanchored ubiquitin chains that interact with the RIG-I:RNA complex to trigger immune signaling [157]. However, both agree that the function of TRIM25 is essential for signal transduction. There is a splice variant of RIG-I, lacking amino acids 36-80, which cannot be ubiquitinated by TRIM25 (or cannot bind poly-ubiquitin) and thus is not capable to transfer a signal [46]. In MDA5 there is no equivalent residue for ubiquitination present [44].

RNF135/Riplet is a second E3 ligase for RIG-I which activates RIG-I by conjugating ubiquitin via K63-linkage to RIG-I. Riplet/RNF135 consists of an N-terminal RING finger domain, C-terminal SPRY and PRY motifs, and shows sequence similarity to TRIM25. Controversial publications propose that it links ubiquitin chains to the CTD of RIG-I [100] or, as RIG-I E3 ubiquitin ligase (REUL), stimulates RIG-I by ubiquitination of the CARDs [48].

One more E3 ubiquitin ligase, RNF125, ubiquitinates the first CARD of RIG-I to mark the protein for proteasomal degradation [4]. The deubiquitinase CYLD [42] negatively regulates RIG-I by removing K63-linked polyubiquitin chains. The same enzyme also acts on CARDIF and further downstream in the signaling pathway [42]. There are a number of ubiquitin like modifiers (UBLs) which have similar functions as ubiquitin, for example small ubiquitin-like modifier (SUMO) or Interferon-induced protein (ISG). RIG-I can additionally be modified by ISGylation [158]. Phosphorylation sites on T170

and S8 further regulate the mechanism by inhibiting (phosphorylated) or promoting (not phosphorylated) ubiquitination of the CARDs by TRIM25 [47, 95].

1.2.6 Inhibitors of the RIG-I like receptors

The mitochondrial protein NLRX1 inhibits the adapter CARDIF [90]. In addition, there are several viral mechanisms to circumvent the immune signaling of the RLR pathway. These include the processing of 5' termini of RNA by dephosphorylation, masking the 5'-PPP [54, 33], cleavage of CARDIF by NS3/4A protein of hepatitis C virus [79] or direct ihibition of the RLRs by by viral proteins. Protein V of paramyxovirus (PIV5-V) interacts with residues 701-830 of MDA5 to suppress its function [21, 104]. G protein of human metapneumovirus [7] and Z protein of new world arenaviruses [39] inhibit RIG-I as well as the NS2 protein of human respiratory syncytial virus (RSV) which binds to the CARDs of RIG-I [82]. The NS1 protein of influenza virus targets the coiled-coil region of TRIM25 to prevent ubiquitination of RIG-I. [53, 45].

1.2.7 Structural studies of the RLR pathway

The structures of the C-terminal domain (CTD) of RIG-I (2RMJ, 2QFB, 2QFD; PDB database accession numbers) [138, 28], MDA5 (2RQB, 3GA3) [138, 77] and LGP2 (2RQA, 2W4R) [138, 77, 107] were determined by NMR and/or X-ray crystallography. RIG-I and LGP2 were co-crystallized as a dimer with a synthetic RNA double stand (3NCU [146], 3LRN, 3LRR[83] , 3EQT, [78]). All structures indicate that this domain has an important function in RNA recognition. In particular, in RIG-I it interacts with the 5'-PPP of target RNA via conserved basic residues. Electron microscopy structures of the receptor LGP2 [91] and of RIG-I as a monomer and dimer [113] were published. The structure of a fragment (292-490) of the helicase domain of MDA5 is deposited in the PDB database under the accession code 3B6E. The structure of the C-terminal single CARD domain of the mitochondrial adapter of RIG-I MAVS has been solved by X-ray crystallography as a maltose binding protein (MBP) fusion protein [110].

1.3 The ubiquitin ligase TRIM25

1.3.1 The superfamily of TRIM proteins

TRIM (tripartite motif-containing) proteins are involved in a large variety of biological processes associated with innate immunity but also genetic disorders, neurological disorders and cancer. The superfamily of TRIM proteins is defined based on the "RBCC" (RING, B-box, coiled-coil) domain arrangement. The RING (really interesting new gene) domain contains a C3HC4 zinc finger of 40-60 residues which binds two zinc atoms. This domain is responsible for protein-protein interactions but has E3 ubiquitin ligase or SUMOylation activity in some cases. The RING domain is followed by one or two B-box domains, of which the second is always present. The B-boxes are CHC3H2 zinc finger motifs, each of about 40 residues, which are exclusively found in TRIM proteins. The B-boxes are characteristically followed by a coiled-coil domain of variable length. This coiled-coil mediates homo- and hetero-oligomeric interactions. The C-terminal region of all TRIM proteins possesses one or two additional functional domains, which classify the TRIM proteins in eleven families. (see Ozato *et al.* 2008 [101] or Nisole *et al.*, 2005 [94] for review). In a large number of TRIM proteins (60% of human TRIMs) the PRYSPRY domain is found as C-terminal motif.

1.3.2 The PRYSPRY domain

The PRYSPRY domain (or B30.2 domain) is present in four protein families. It is found as C-terminal domain in the majority of all TRIMs, in butyrophilin-related transmembrane glycoproteins, in stonustoxin (a cytosolic secreted protein) and in the entherophilin-related family. The SPRY domain alone is found in the N-terminal region of single strand binding (SSB) proteins that belong to a family of suppressors of cytokine signaling (SOCS). The PRYSPRY domain is assembled of two parts: the about 60 amino acid PRY segment is followed by the SPRY element of approximatively 140 residues. The SPRY domain is more ancient than the PRYSPRY domain and can occur alone, but in 53% of cases it has fused with the PRY domain [50, 116].

Characteristically, the PRYSPRY domain has an immunoglobulin-like fold, comprising two β-sheets which form a hydrophobic sandwich core connected by hypervariable loops that determine the surface shape properties of the domain. There are several high resolution structures of PRYSPRY domains available, including the PRYSPRY of human sRFPL1 (2FBE) [50], a protein of unknown function, *Drosophila melanogaster* GUSTAVUS in complex with peptides of its interaction partners VASA and elongin B and C [151, 150] (2IHS with VASA peptide, 2FNJ with elongin B and C), the ones of human and *murine* TRIM21 alone and in complex with an Ig Fc fragment [61, 70] (human: 2IWG, murine: 2VOK, 2VOL), the PRYSPRY domain of human TRIM72 (3KB5), the one of human pyrin (2WL1) [148] and several SPRY domains of SOCS box proteins.

Among the family of PRYSPRY containing TRIM proteins there are proteins with direct medical impact, like pyrin (TRIM20), TRIM5α or MID1 (TRIM18). Mutations in pyrin cause the autoinflammatory Familial Mediterranean Fever syndrome (FMF) [148]. The exact function of pyrin is unknown but it seems to be involved in interleukin-1β maturation and secretion. TRIM5α is a cytosolic retroviral restriction factor that blocks retroviruses after cell entry and before integration [128]. In some monkey species TRIM5α completely inhibits HIV-1, whereas in humans it only modestly blocks HIV-1 but efficiently inhibits N-tropic murine leukemia virus (N-MLV). Mutations in the PRYSPRY domain of MID1 (TRIM18) cause the Opitz syndrome, characterized by multiple abnormalities of midline structures due to defects in the embryonic development at the midline [31]. A sequence alignment of several SRPY domains is shwon in figure 1.10, variable loop regions and residues important in disease or mutagenesis studies are annotated.

By means of the PRYSRPY complex structures and mutational analyses, critical residues for protein-protein interaction in the PRYSPRY domains have been identified and mapped onto the protein surface. Peptides with different conformations and se-

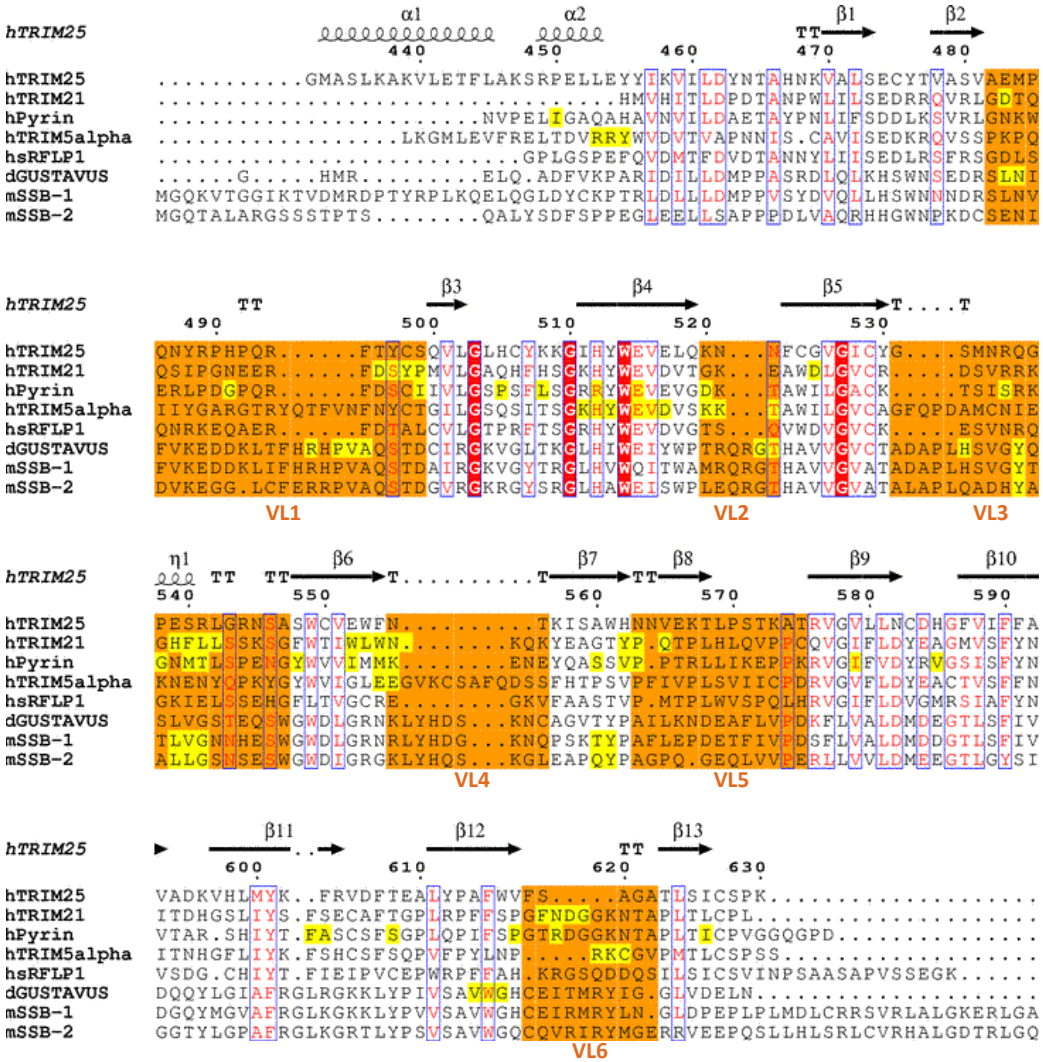


Figure 1.10: *Alignment of PRYSPRY domains of which complex structures or relevant mutants are known. On the top the secondary structure elements associated with TRIM25 are shown. The variable loop regions are marked in orange. In yellow there are highlighted: in TRIM21:Fc contact residues; in Pyrin: residues when mutated cause FMF; in Trim5α: residues where a triple alanine constitution disrupts retroviral restriction activity; GUSTAVUS: residues which interact with the VASA peptide; mSSB-1 and mSSB-2: mutations which affect Par-4 binding. Conserved regions are indicated by red color and blue boxes. Secondary structure of TRIM25 is mapped on top of the alignment, the TRIM25 sequence represents the crystallized construct.*

quences can bind to the different PRYSPRY domains. The form of interaction pockets and thus the mode of interaction vary greatly among different PRYSPRY domains [148]. This and the fact that this domain has evolved in a variety of cellular contexts and functionally unrelated proteins indicate that the fold serves as a scaffold for protein-protein interactions. The structures of GUSTAVUS bound to a VASA peptide and peptides of elongin B and C on different sides of the molecule indicates that also other PRYSPRY containing proteins might be able to interact with several partners at the same time via different interaction surfaces [151, 150].

1.4 Influenza virus

1.4.1 Influenza

Influenza or commonly “flu”, is an infectious disease caused by viruses of the *Orthomyxoviridae* family. Common symptoms of influenza are chills, fever, sore throat, muscle pains, headache, coughing, weakness, fatigue and general discomfort. Typically, influenza in humans is transmitted in aerosols containing the virus (sneezing) or by direct contact with bird droppings or nasal secretions of an infected person. There are three classes of influenza, influenza A, B, and C. The influenza B virus infects only humans and seals, whereas influenza C virus causes mild pathogenic symptoms in human and pig. Influenza A viruses can infect birds and mammals and are further classified into subtypes based on the antigenicity of their surface molecules hemagglutinin (HA) and neuraminidase (NA). There are 16 subtypes of HA (H1-H16) and nine subtypes of NA (N1-N9). The subtypes H1N1, H1N2, and H3N2 are infectious for humans. The nature of HA determines the host-specificity since different subtypes of HA bind to different N-acetyl neuraminic acids (or sialic acid (SA)), that differ between host species. Influenza virus evolve extremely fast due to a lacking proofreading activity of the polymerase (on average one point mutation per genome copy) (antigenic drift) and the segmented genome type, which enables genomic recombination in cells infected by multiple viruses (antigenic shift) [103].

1.4.2 Structure of the influenza virus particle

The Influenza A virion contains a segmented genome of eight negative sense single RNA strands and the segments code for eleven proteins: i. e. HA, NA, NP, M1 and M2 (same segment, different reading frames), NS1 and NEP, PA, PB1 and PB1-F2, PB2. The total genome size is about 13,600 bases. The virus particles have an average diameter of about 100 nm. The virus consists of an outer lipid bilayer envelope in which the glycoproteins HA and NA (in a ratio of about 4:1) and the transmembrane protein M2 (one M2 per 10^1 to 10^2 HA) are inserted. Inside the outer membrane there is an inner protein layer

formed by the matrix protein M1. The core contains the nuclear export protein NEP (also called nonstructural protein NS2) and the ribonucleoprotein (RNP) complex consisting of the genomic RNA coated with the nucleoprotein (NP) and the heterotrimeric viral RNA polymerase complex (PB1, PB2 and PA) (figure 1.11). The ends of the RNA segments are highly conserved among all influenza viruses and form a helical hairpin ("panhandle structure") bound to the polymerase complex.

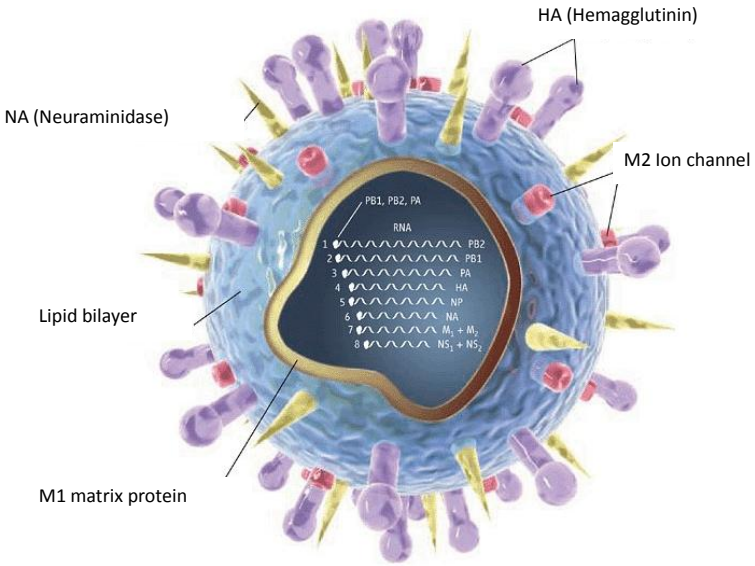


Figure 1.11: *The structure of the influenza virus. [5]*

1.4.3 Influenza virus replication

When infecting a cell, the surface protein HA interacts with sialic acid (SA) on the target cell surface. Sialic acids are nine-carbon acidic monosaccarides found at the termini of glycoconjugates at the cell surface. After attachment, the virus is internalized by endocytosis in acidic vesicles. The low pH in the vesicles is crucial to the virus as it induces a conformational change in HA. This change initiates fusion of the viral and cellular membranes, allowing the virus to enter the cytoplasm. Secondly, the proton channel M2 pumps protons into the interior of the virus particle, triggering the uncoating and the release of the ribonucleoprotein particles. The ribonucleoprotein particles are transported to the nucleus where transcription of the viral RNA by the viral RNA polymerase takes place.

Export of the ribonucleoprotein particles to the cytosol is mediated by NEP and M1. Translation takes place on membrane-bound ribosomes into the endoplasmatic reticulum, where the proteins are folded and transferred to the Golgi apparatus for post-translational modification. New virus particles assemble at the plasma membrane of the host cell, that has been prepared for viral budding by insertion of the viral membrane proteins HA, NA and M2 (see Arias *et al.*. and Bouvier *et al.* for review [3, 14]).

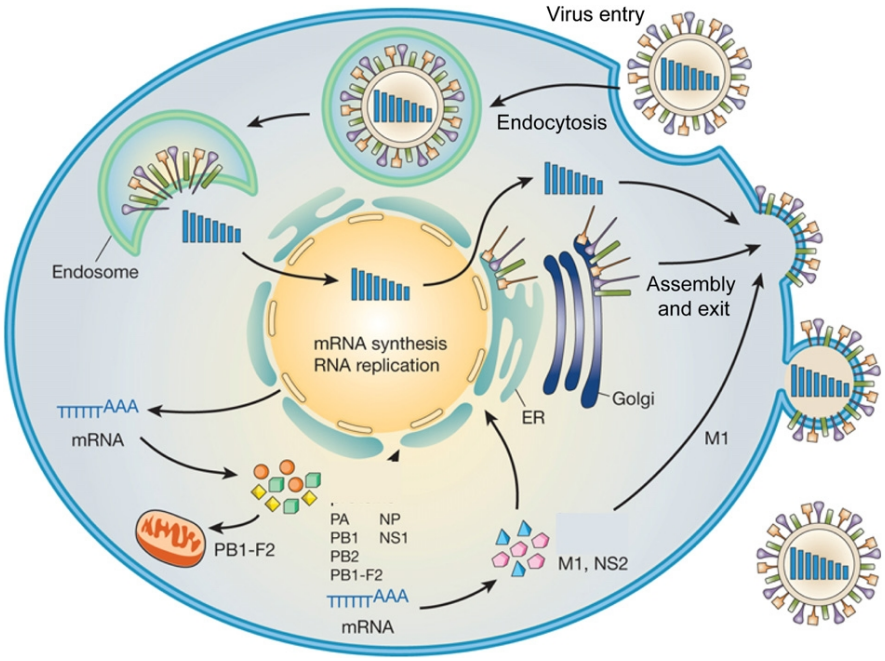


Figure 1.12: *The replication cycle of the influenza virus [3]*

1.4.4 Vaccination and Medication

Vaccination

Infection by influenza viruses can be prevented by vaccination. However due to the fast evolving virus, annual re-vaccination of individuals with a constantly updated vaccine is required. The trivalent seasonal influenza vaccine consists of HA surface glycoprotein components from the influenza subtypes H3N2, H1N1 and from influenza B viruses.

M2 channel inhibitors

The M2 ion channel is a tetrameric integral membrane protein. It is targeted by drugs of the adamantanes class (adamantanes, amantadine, rimantadine), which block the ion channel activity. This class of drugs is efficient against all Influenza A viruses but not against Influenza B or C. However, resistance against these kind of drugs increases.

Neuraminidase inhibitors

The most known drug of the family of neuraminidase inhibitors is oseltamivir ("Tamiflu"). Oseltamivir and zanamivir target the NA of influenza viruses and are active against influenza A and B viruses. NA is a homotetrameric glycoprotein anchored in the viral membrane catalyzing SA removal from its linkage to galactosidase. Blocking this reaction suppressed the release of viral particles. Resistance against these drugs is emerging.

Further Alternatives

There is a need for the development of new anti-influenza therapies as the recent 2009 influenza pandemic has demonstrated. Several new compounds are already in the pipelines of the pharmaceutical industries; for example the nucleoside analogs T-705, Ribavirin and Viramidine, that inhibit polymerase activity. Anti-sense RNA therapeutics targeting the influenza virus RNA have shown efficacy in model organisms. Peramivir and CS-8958 (pro-drug of R-125489) are NA inhibitors structurally different from the substrate SA (unlike zanamivir and oseltamivir). Recently, a high affinity neutralizing antibody against HA ectodomain has been shown to act efficiently against a row of different influenza viruses. For a recent review on anti-viral drugs see Das *et al.* [30].

The influenza virus polymerase as a novel drug target

The influenza virus RNA-dependent RNA polymerase fulfills essential functions for the virus. It uses negative-sense viral RNA as a template to synthesize positive-sense RNA species: i. e. mRNA templates for protein synthesis and complementary cRNA inter-

mediates from which subsequently new copies of the negative-sense genomic RNA are transcribed. The polyadenylation (poly(A)) tail of viral mRNAs is encoded in the viral genome. Capping of the messenger RNA is achieved by a mechanism unique for influenza virus, bunyaviruses and arenaviruses, in which the polymerase ”steals” 5’-capped primers from cellular pre-mRNA transcripts to initiate protein synthesis. This mechanism is called ”cap-snatching”. Since function of the viral RNA polymerase is indispensable for the virus, inhibition of its function by either suppressing cap-binding, endonuclease or polymerase activity could be the basis for new therapies [10, 35, 52].

Further introductory information is provided in the article manuscript included in this thesis in chapter 4.

1.5 Aim of the thesis

The RLR pathway is a relatively newly discovered immune signaling mechanism [154]. To date, the CTDs of all three receptors have been characterized in atomic resolution [28, 138, 77, 107]. This part of the protein has an important function RNA interaction and is the site of 5'-PPP recognition, but it represents only a small part of the whole molecule (15%). Much further work is necessary to completely understand the mechanisms and regulation of the signalling pathway and how viral counter measures inhibit it. Structural investigation should focus on:

- The specific recognition of viral RNAs by the receptors, the actual contribution of the helicase domain to RNA binding (especially for MDA5), the contribution of the ATPase function to the mechanism and the conformational rearrangements governing the interactions between the molecules and the interaction with the adapter molecule MAVS.
- The characterization of the mechanisms underlying the regulation of RIG-I by poly-ubiquitin chains and the E3 ubiquitin ligase TRIM25.
- Understanding the mechanisms of viral proteins which inhibit signaling on the level of the RLRs.

The major goal of this thesis work was to obtain a high resolution crystal structure of RIG-I and its complex with RNA. Constructs of RIG-I were recombinantely expressed and purified, and RNA ligands were produced. Complexes of RIG-I with different RNAs were formed and characterized but failed to crystallize. However, low resolution small angle scattering (SAXS and SANS) structures of RIG-I alone and in complex with RNA were modeled. We investigated interaction of RIG-I with it's E3 ubiquitinase TRIM25. Furthermore the crystal structure of the PRYSPRY domain of TRIM25 is presented. There will be a summary of the work done so far on the recombinant expression and purification of the complex of MDA5 with its viral inhibitor V protein of paramyxovirus.

In addition, a draft manuscript is included in this thesis, investigating the co-crystal structures of the A/California/04/2009-H1N1 influenza polymerase endonuclease in complex with four polymerase inhibitors, as well as dTMP and rUMP as a basis for novel anti-influenza drugs. This manuscript is confidential.

Chapter 2

Results

Summary of chapter results

As a basis for structural and biochemical studies of the molecular mechanisms of RIG-I like receptors, several constructs of human RIG-I were recombinantely expressed and protocols were established to produce pure and homogeneous proteins. Moreover, different putative RNA ligands of RIG-I were cloned and produced by T7 transcription. RIG-I:RNA complexes could be assembled in a 1:1 stoichiometry. The reported interaction of RIG-I with polyubiquitin or direct *in vitro* ubiquitination of RIG-I could not be reproduced in our hands. Although human RIG-I and RIG-I:RNA complexes failed to crystallize, small angle scattering data were collected and a low resolution models of the protein and protein:RNA complexes were derived. The PRYSPRY domain of TRIM25, the putative E3 ligase of RIG-I, was crystallized and the structure of this domain was determined at a resolution of 2 Å. Very recently, full length mallard duck RIG-I was crystallized and a complete native data set was collected and the structure will now be solved by experimental phasing. A complex of the receptor MDA5 with its viral inhibitor V protein from parainfluenza virus could be expressed and purified.

Constructs of the endonuclease domain of influenza virus A/California/04/2009-H1N1 were expressed in *E.coli*, purified and crystallized alone and with four diketo acid inhibitors, dTMP and rUMP. All ligands bind to the two manganese ions in the active site of the protein, as well as to a number of protein residues. Some residues change conformation upon substrate/inhibitor binding. The structures of these complexes provide important information for design and optimization of new anti-influenza drugs.

Résumé du chapître "Résultats"

Dans le cadre des études structurales et biochimiques des voies d'activation des hélicases de type "RIG-I", différentes constructions de RIG-I ont étés produites et un protocole de purification a été établi permettant l'obtention d'une protéine pure et homogène. Malgré le fait que nous ne sommes pas arrivés à cristalliser la protéine RIG-I, des données de SAXS et de SANS ont permis de construire une structure basse résolution. Récemment, nous avons réussi à obtenir un jeu de données de diffraction provenant d'un RIG-I de canard mais la structure n'a pas pu être résolue. Nous avons également obtenu par transcription *in vitro* plusieurs ARN qui forment un complexe avec RIG-I. De plus, nous n'avons pas réussi à reproduire de précédents résultats montrant une interaction in vitro entre l'ubiquitine et RIG-I ou encore son ubiquitination. Néanmoins, le domaine PRYSPRY de la protéine TRIM25 qui possède la fonction d'ubiquitine ligase E3 de RIG-I a pu être cristallisé. La structure de ce domaine a été obtenu à 2 Å de résolution. Un complexe comprenant le récepteur MDA5 ainsi que l'inhibiteur "protéine V" provenant du virus parainfluenza a pu être établi.

La construction du domaine de l'endonucléase du virus de la grippe du porc H1N1 a été exprimée, purifiée et cristallisée avec des inhibiteur (acide diketo) ou en présence de nucléotides libres (TMP ou rUMP). Les inhibiteurs et les nucléotides fixent les deux ions manganèses présents dans le site actif de la protéine ainsi que plusieurs autres acides aminés. Certains résidus subissent un changement de conformation après la fixation des substrats ou des inhibiteurs. La structure des ces complexes permet de fournir d'importantes informations concernant la synthèse et l'optimization de nouvelles molécules antivirales spécifiques du virus de la grippe.

2.1 RIG-I

2.1.1 Protein Expression

Human RIG-I

After testing of various constructs, with and without solubility improving tags (His, GST, MBP, Smt3) and expression in different media (LB, TB, auto inducible medium [135]) only RIG-I CARDs (constructs 1-226 and 1-208) could be expressed and purified from *E.coli*. All other constructs were expressed in insect cells. A standard purification protocol included a first nickel affinity column, TEV cleavage and dialysis, a second nickel affinity column and size exclusion chromatography as a polishing step (details are given in the Material and Methods, chapter 5). From a pool of divers fragments, found by rational design based on alignments, and on results of tryptic digests, the following were used in the study: RIG-I full length (1-925), RIG-IΔCARDS (239-925), RIG-IΔCTD (1-793), RIG-I helicase (293-793) (figure 2.1). All constructs could be purified to homogeneity (figures 2.2, 2.3, 2.4, 2.5, 2.6).

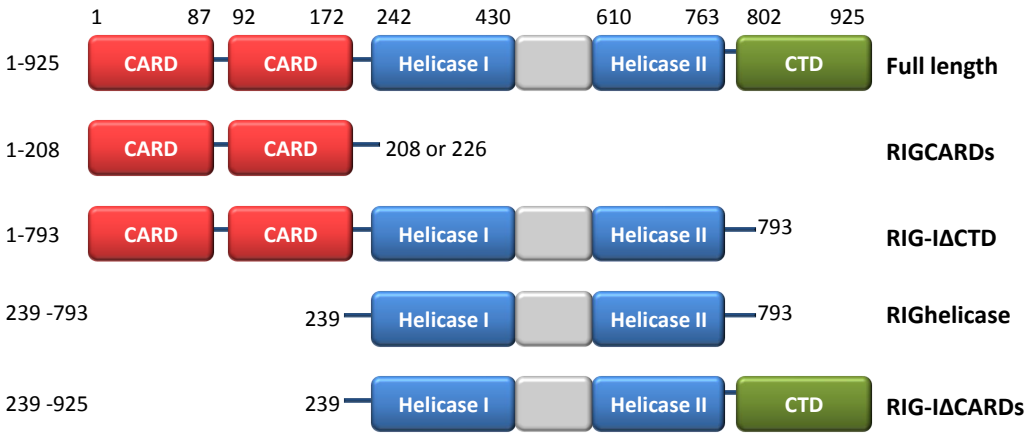


Figure 2.1: *Human RIG-I constructs used in the study.*

Other species

Since human RIG-I constructs were nicely behaved but failed to crystallize, homologous proteins from other species was tested, since even small differences in the amino acid sequence can alter protein properties to enable crystallization. Expression trials of RIG-I constructs of *Ornithorhynchus anatinus* (Platypus) were unsuccessful. Full length *Ornithorhynchus anatinus* RIG-I (1-910) and the fragments 136-910, 227-910, 227-778 and 1-778 could be expressed in insect cell culture but were insoluble. However, RIG-I from mallard duck (*Anas platyrhynchos*) [8] was expressed in insect cells, could be purified to homogeneity and crystallized (see section 2.1.7).

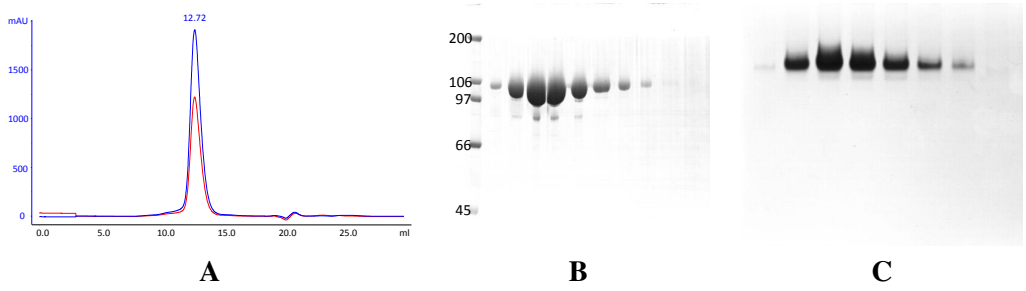


Figure 2.2: *Purification of human full length RIG-I. (**A**) Size exclusion chromatography profile on S200/10/300 column. (**B**) SDS-PAGE of the peak fractions. (**C**) Native PAGE of the peak fractions.*

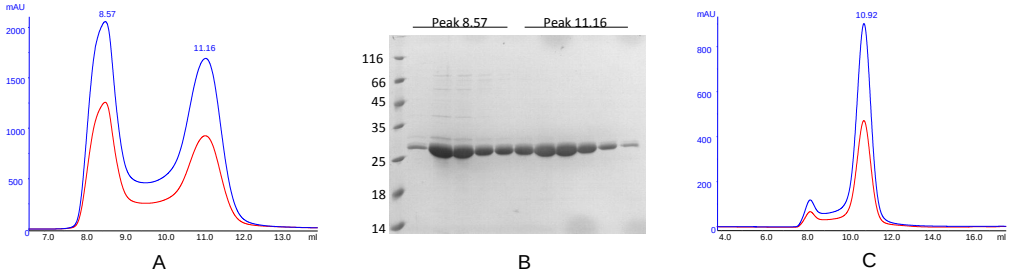


Figure 2.3: *Purification of human RIG-I CARDs (residues 1-208) **(A)** Size exclusion chromatography profile on S75/10/300 column. The peak at 8.5 ml elution volume represents aggregated protein. The peak at 11.16 ml represents the monomeric CARDs. **(B)** The corresponding peak fractions on an SDS gel. **(C)** Re-inject of the concentrated fractions of the non-aggregate peak on the S75/10/300 column. CARDs stay mostly monomeric after concentration.*

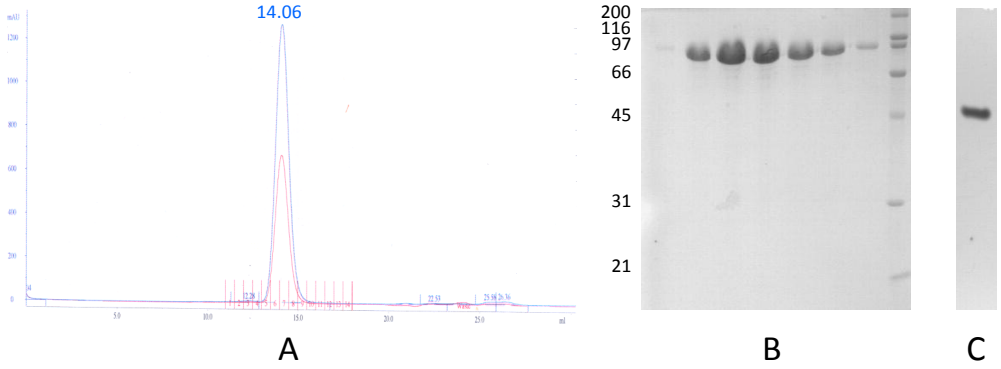


Figure 2.4: *Purification of human RIG-IΔCTD (residues 1-793).* ***(A)*** *Size exclusion chromatography profile on S200/10/300 column.* ***(B)*** *The corresponding peak fractions on an SDS gel.* ***(C)*** *Native PAGE of purified protein indicates homogeneity.*

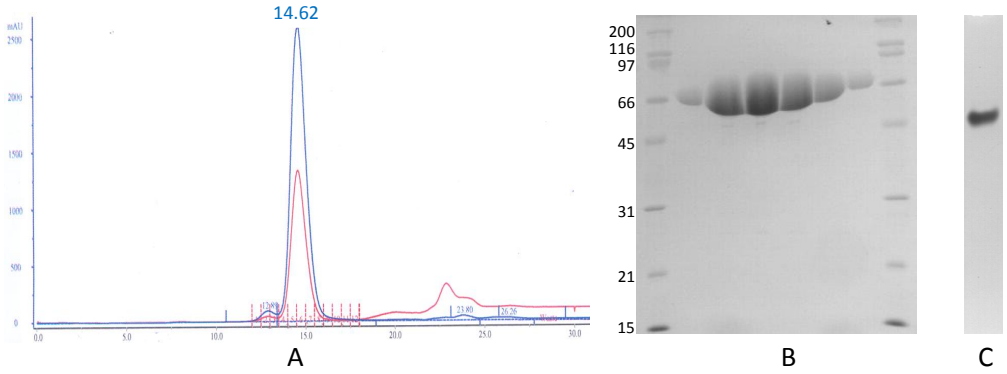


Figure 2.5: *Purification of human RIG-I helicase (residues 239-793).* ***(A)*** *Size exclusion chromatography profile on S200/10/300 column.* ***(B)*** *The corresponding peak fractions on an SDS gel.* ***(C)*** *Native PAGE of purified protein indicates homogeneity.*

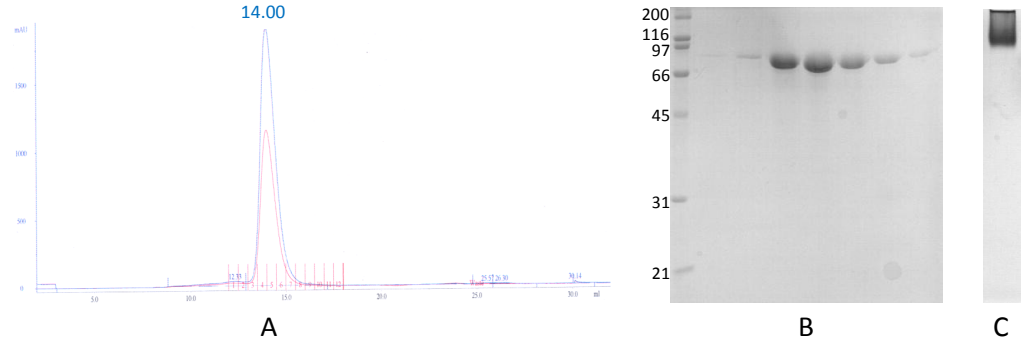


Figure 2.6: *Purification of human RIG-IΔCARDs (residues 239-925)* ***(A)*** *Size exclusion chromatography profile on S200/10/300 column.* ***(B)*** *The corresponding peak fractions on an SDS gel.* ***(C)*** *Native PAGE of purified protein indicates slight polydispersity.*

2.1.2 Activity of RIG-I

Purified full length RIG-I from insect cells was tested for ATPase activity (Figure 2.7). In agreement with published data, RIG-I is an active RNA dependent ATPase, able to hydrolyze ATP exclusively in presence of RNA [92]. The RNA used in this experiment was a synthetic blunt end non-phosphorylated 40 nucleotide perfect double strand of sendai virus leader. Leader RNAs are 5'-regions in the viral genome being non-transcribed, but carrying important signals to control transcription, replication, and packaging of the viral genomes into new virions. The experiment was performed in collaboration with Stephane Haussman (Kolakofsky lab, Geneva).

2.1.3 Thermal stability assay on human RIG-I

Purified full length RIG-I was perfectly pure and homogeneous after size exclusion chromatography but failed to crystallize. To find stabilizing conditions or additives, a thermal shift (or Thermofluor) assay (developed by Pantoliano and coworkers, US patent 6,020,141 [102]) was performed. The higher the melting temperature (T_m) of the protein in the assay, the more stable it is and the higher is the probability of the protein to crystallize [38]. For full length human RIG-I, all additives tested show curves with the T_m in the same range of about 43 °C to 45 °C. Thus no stabilizing effect of the tested additives could be observed (Figure 2.8). The tested additives included ZnCl, $MgCl_2$, $MnCl_2$, ATP + $MgCl_2$, ADP + $MgCl_2$, AMPPCP + $MgCl_2$, a 20nt RNA, Na/KPO_4 and $MgSO_4$.

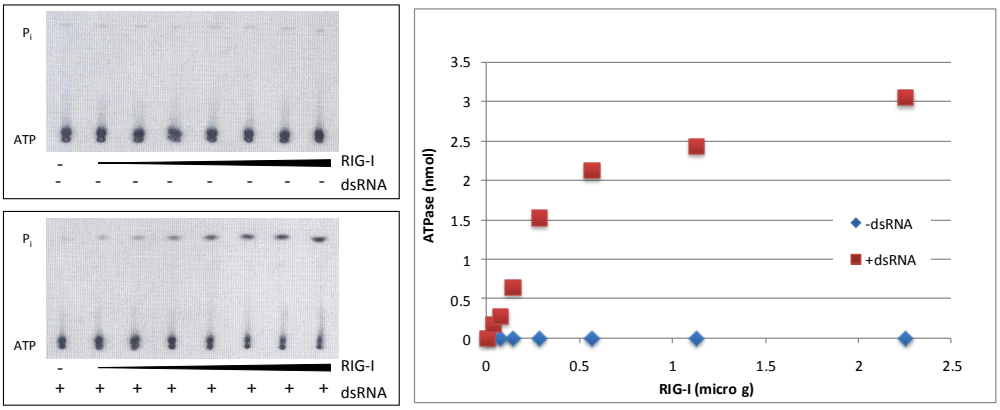


Figure 2.7: *Purified RIG-I has RNA dependent ATP hydrolysis activity.*

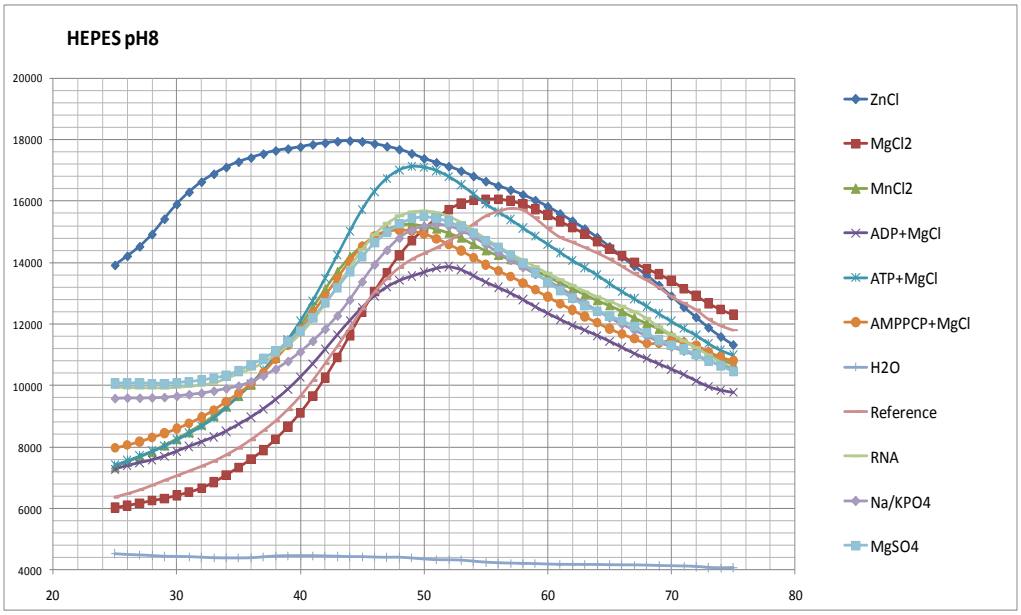


Figure 2.8: *Additives tested by Thermofluor assay did not have any stabilizing effect on full length RIG-I.*

2.1.4 RIG-I RNA interaction

RNA production

As RIG-I interacts with various 5'-PPP-RNAs *in vivo* and *in vitro*, we wanted to establish RIG-I:RNA interaction with the purified protein. A complex with RNA could also be a means to stabilize RIG-I conformationally to facilitate crystallization. For this, RNA constructs which putatively serve as target RNA for RIG-I were chosen. A clone of a short panhandle RNA from influenza virus (shPH, 35 bases) was available in the lab and provided by Thibaut Crépin. A measles leader (MVL, 55 bases) RNA clone which had already shown activity *in vivo* was provided by the lab of Denis Gerlier, Lyon [108]. Furthermore, rabies virus leader (RVL, 61 bases) RNA was cloned, as it had shown to interact with RIG-I *in vitro* [28] and a construct of ebola virus leader (EVL, 58 bases) RNA was designed. Leader RNAs comprise the 5' region of genomic viral RNA and are not transcribed but carry regulatory information. The sequences of these RNAs with their secondary structure predictions by MFOLD [160] are shown in figure 2.9. The cloned sequences contained the T7 promoter followed by the desired sequence, preferably starting with a G at the 5'end. An enzymatic restriction site for linearization of the plasmid was included at the 3'end . For RNA production, T7 run-off transcription on a linearized plasmid was performed. After, the reaction was loaded onto a denaturing urea-PAGE gel, the band of desired size was excised and the RNA was extracted from the gel. Gels of several typical purifications are shown in figure 2.10. Differences between unpurified transcription mixes and the purified products are obvious (figure 2.10 A and B). However, transcription products differing by only several bases cannot be separated. Native PAGE of the RNAs shows homogeneity of the EVL and shPH RNAs (figure 2.10 C). The RVL construct might contain a larger species. Weather the RNA is folded like predicted from MFOLD cannot be controlled during the purification. In later experiments, aliquots of the *Alu* domain of *Pyrococcus horikoshii* RNA from signal recognition particle (SRP) RNA and a *Candida alibicans* $tRNA^{Asn}$, purified in the same way, were provided by Thibaut Crepin.

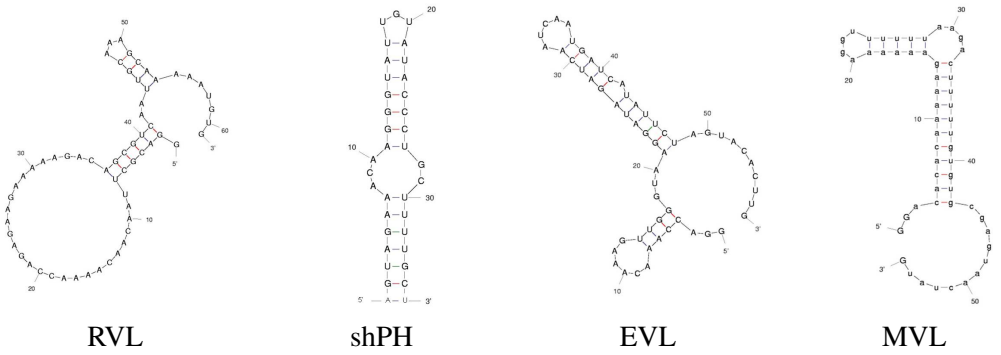


Figure 2.9: *RNA constructs used in this study represented in a minimal energy structure as calculated by MFOLD [160]. Rabies virus leader (RVL), influenza short panhandle (shPH), Ebola virus leader (EVL) and Measles virus leader (MVL).*

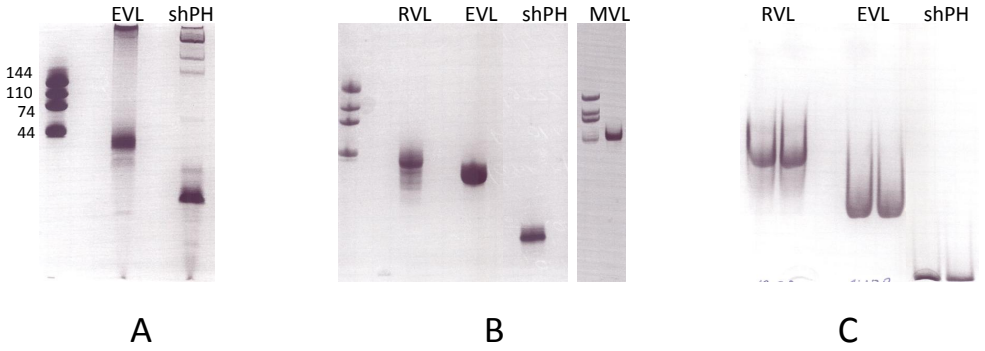


Figure 2.10: *Typical gel profiles of RNA purifications. (**A**) Non-purified T7 transcription of EVL and shPH RNA. (**B**) Purified RNA after gel elution and desalting. RVL, EVL, shPH and MVL are shown. The RVL RNA has degradation by-products. (**C**) Native TBE gels of purified RNA. EVL and shPH show one species on the gel, while a second species could be suspected for RVL.*

Interaction with different RNA species

The RIG-I:RNA interaction was tested with two methods: on native gels and by multi angle laser light scattering coupled to a size exclusion column (SEC-MALLS). On native gels, a band shift indicating the protein:RNA complex can be observed (figure 2.11). In the titration experiment, unbound RNA is present when the molar ratio of RIG-I:RNA exceeds 1:1, which indicates a binding of one molecule of RIG-I to one RNA molecule. In SEC-MALLS experiments, shape and size of macromolecules are analyzed, dependent

on their elution volume from the size exclusion column in combination with the apparent molecular mass from light scattering.

Complex formation between RIG-I and different RNA species can be observed by a peak shift in size exclusion chromatography on a S200/10/300 column for shPH, EVL, MVL and RVL (figure 2.12). The shift to a lower volume indicates a larger, more elongated or less globular particle. This can point to a conformational change of the protein molecule, the addition of the RNA moiety to one end of the protein elongating the whole complex, or that bound RNA is attached but large parts of it are flexible and floppy. In the shown chromatograms, the unbound RNAs elute as single peaks, indicating monomeric RNAs relatively homogeneous in size. However, when looking at the apparent molecular masses from MALLS, not all complexes show a significant mass shift (figure 2.12). RIG-I alone appears with a mass of 100 kDa, slightly smaller than the calculated mass of 106 kDa, but within the error range. In MALLS, a deviation from the calculated mass can be normal for non-globular proteins, since the apparatus (light scattering detector coupled to a refractive index detector) is calibrated with the globular shaped protein bovine serum albumin (BSA). This fact explains also the high error rate for the free RNA masses, as RNA scattering and refraction properties are different from protein. Indeed, the SAXS structure of RIG-I reveals an elongated shape of the molecule (see section 2.1.5). The shPH RNA and RVL RNA in complex with RIG-I show a MALLS mass shift of around 10 kDa, which can be accepted as significant and appropriate for the mass of the RNA. EVL and MVL do not induce a significant mass shift, however the resulting complex is monodisperse (one peak, flat MALLS signal). From the SEC-MALLS comparison of EVL, MVL, RVL and shPH, it can be concluded that RIG-I binds all these RNAs and forms a homogeneous complex, probably with a 1:1 stoichiometry. RVL and shPH RNA induce significant mass shifts in MALLS which suggest a 1:1 binding ratio. For this reason we focused on those latter RNAs for most of the following experiments.

But does RIG-I also bind cellular non-target RNAs? SRP *Alu* RNA and a tRNA which are abundant in the cell but not natural ligands for RIG-I were tested. In a MALLS experiment, interaction of RIG-I with a 110 kDa *Alu* SRP RNA from *Pyrococcus horikoshii* is observed (figure 2.14 A). However, RIG-I might not interact with cellular SRP RNA for two reasons. Firstly, human SRP RNA is tightly bound to the proteins of the SRP particle an the 5'-PPP is masked. Secondly, the 5' and 3' end of the archeal *Alu* domain form a helical structure with a blunt end and thus would be a ligand for RIG-I, while the human version does not have complementary ends [147]. The tested tRNA from *Candida alibicans* does not interact with RIG-I (figure 2.14 B). tRNAs have 3' and 5' ends in a stem structure with a 3' overhang of four bases where the amino acid is attached. RIG-I seems to bind preferentially RNAs with blunt ends or, 5' overhangs with lower affinity. It is controversially discussed if it interacts with 3' overhangs or not [123, 96].

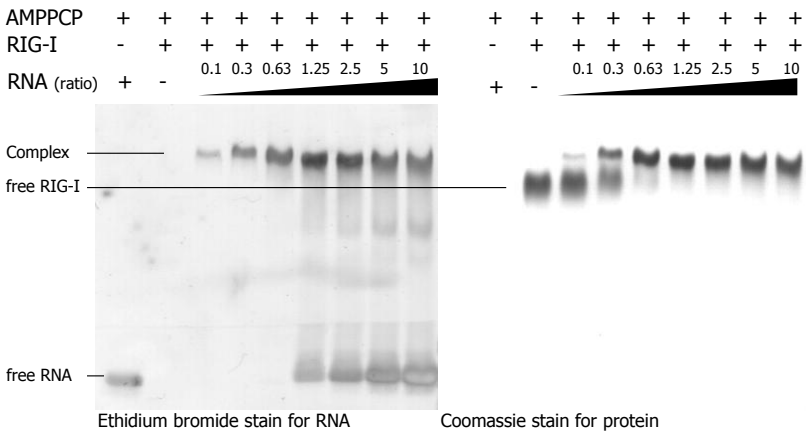


Figure 2.11: *RNA (shPH) titration on RIG-I on native TBA gel. Unbound RNA can be observed starting from a molar ratio higher than 1:1.*

ATP dependence

The finding that full length RIG-I forms 1:1 complexes with the various tested RNA is contradictory to studies that propose dimerization or multimerization upon RNA binding [28, 77, 125, 121, 113]. This will be further discussed in the Discussion, chapter 3. All

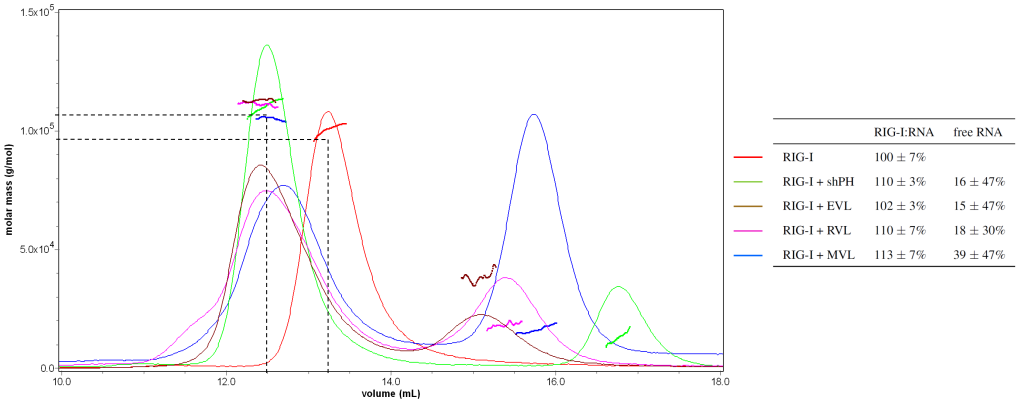


Figure 2.12: *SEC-MALLS of full length RIG-I with different RNA species.*

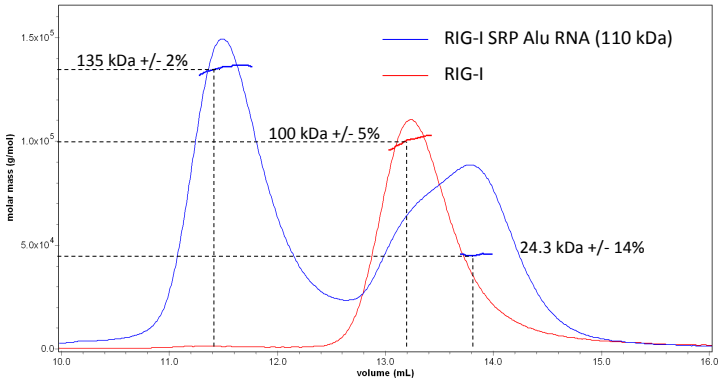


Figure 2.13: *SEC-MALLS results of full length RIG-I and Pyrococcus horikoshii SRP Alu RNA.*

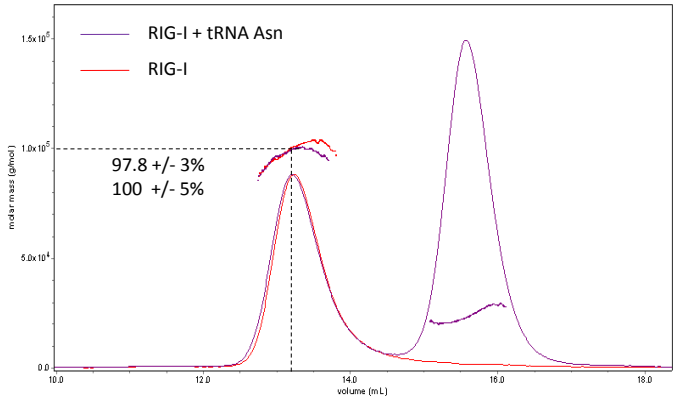


Figure 2.14: *SEC-MALLS results of full length RIG-I and Candida alibicans tRNA^{Asn} RNA.*

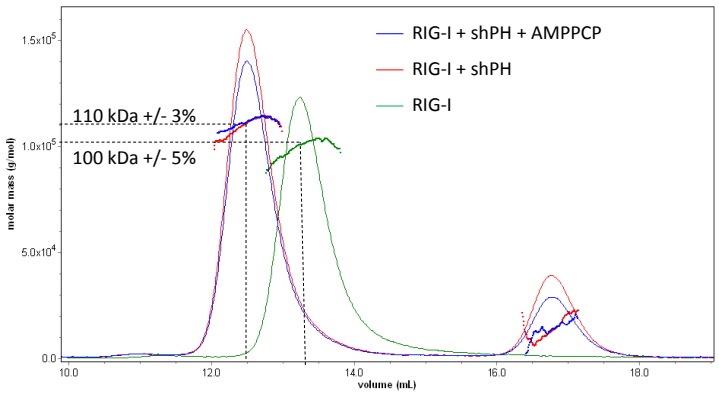


Figure 2.15: *SEC MALLS curves of full length RIG-I with shPH RNA and AMPPCP.*

RNA binding experiments were so far performed ignoring the fact that RIG-I has ATPase activity. To test weather ATP has an influence on RNA binding and oligomeric state, a SEC-MALLS experiment of RIG-I with RNA in presence of the non-hydrolysable ATP analogue AMPPCP was performed. (figure 2.15). With ATP analogue the complex retains a 1:1 binding ratio.

Domains responsible for RNA binding

To determine which domains of RIG-I are responsible for RNA binding, the interaction of C- and N-terminal truncation mutants of RIG-I with RNA was tested. RIG-IΔCTD lacks the CTD, RIG-IΔCARDs lacks the CARDs and the RIG-Ihelicase lacks both, CARDs and CTD. The purified protein fragments were incubated with shPH RNA and subjected to native PAGE analysis (figure 2.16). Full length RIG-I and RIG-IΔCARDs show a gel shift on the native gel, while RIG-IΔCTD does not. These results were confirmed by SEC-MALLS analysis in which RIG-IΔCTD and RIG-Ihelicase (both lacking the CTD) do not interact with RNA (figures 2.17 and 2.18). The SEC-MALLS result for RIG-IΔCARDs is surprising, as a shift of the peak in the chromatogram goes in hand with a constant value for the mass (figure 2.19). This phenomenon could be a result of either a conformational change upon RNA binding or a less floppy RNA which is more tightly bound than in the case of the full length protein. For similar constructs of RIG-I lacking the tandem CARDs, elevated ATPase activity and 15-fold accelerated movement along dsRNA was

observed compared to the wild type [92, 28]. Possible conformational changes of this mutant have to be further investigated by SAXS. From the analysis of the RIG-I mutants, it can be concluded that the CARD domains are not necessary but the CTD is needed for RNA binding.

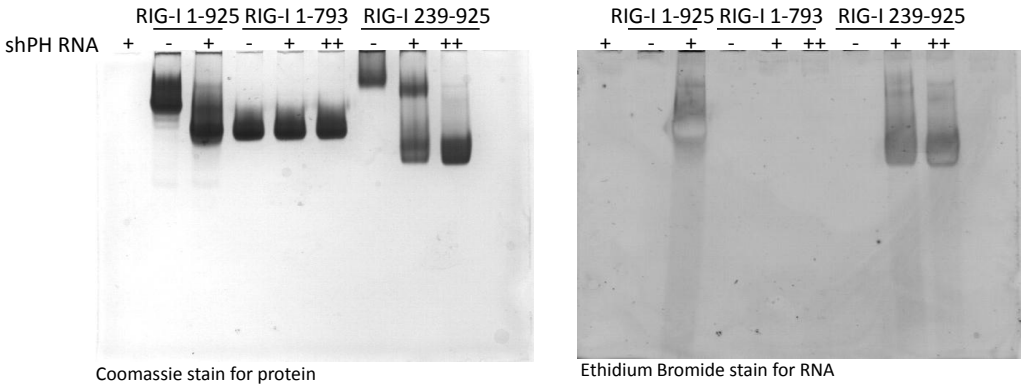


Figure 2.16: *RNA (shPH) binding of RIG-I truncations (Tris/glycine native gel). While full length RIG-I binds to RNA, RIG-IΔCTD (1-793) does not. RIG-IΔCARDs binds to RNA.*

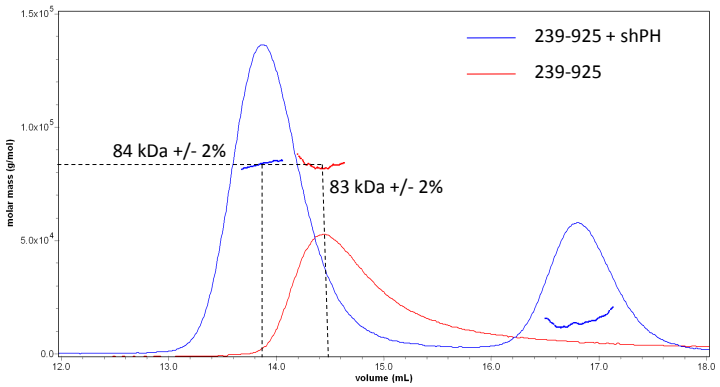


Figure 2.17: *MALLS data of RIG-IΔCARDs in complex with shPH RNA.*

Conformational changes upon RNA binding

To reveal conformational differences of RIG-I in presence of RNA, a trypsination experiment was performed with the full length protein. RIG-I alone, RIG-I with shPH RNA (1:2 ratio) and RIG-I with shPH RNA and AMPPCP were incubated with trypsin in a 1:500 ratio. At several time points the reaction was stopped and the samples were analyzed

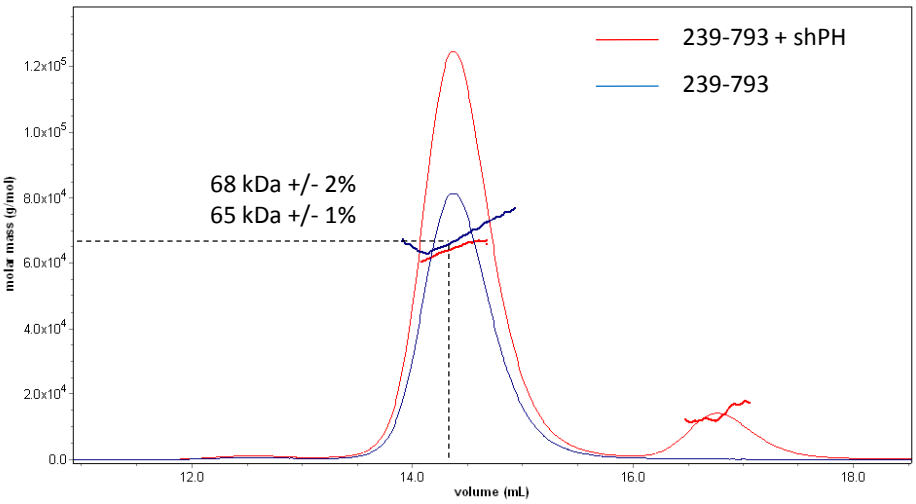


Figure 2.18: *MALLS data of RIG-I helicase in complex with shPH RNA.*

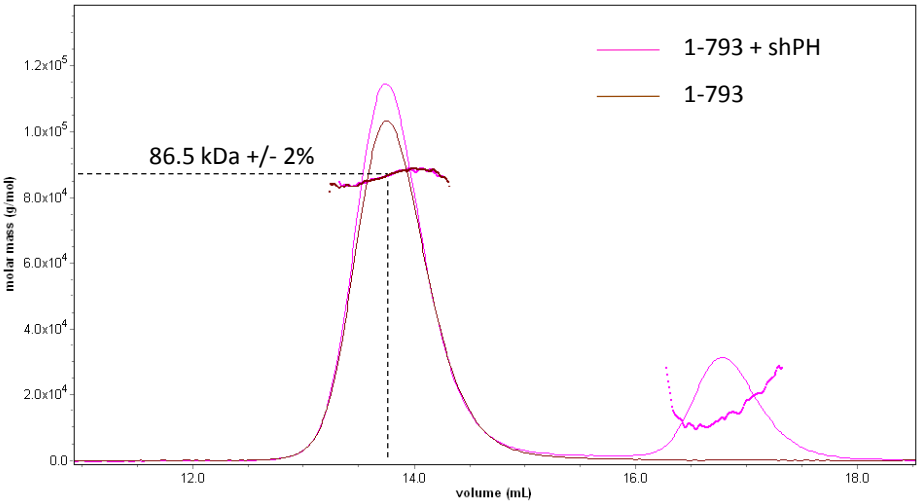


Figure 2.19: *MALLS data of RIG-IΔCTD in complex with shPH RNA.*

by SDS PAGE (figure 2.20). An apparent change in digestion pattern can be observed between the protein alone and the samples containing RNA. Two additional bands (red boxes) appear for the RIG-I:RNA complex. These bands were analyzed by mass spectroscopy and N-terminal sequencing and comprise residues 680-925 for the 30 kDa band and residues 730-925 for the 20 kDa band. In agreement with the literature, it can be concluded that upon RNA binding a fragment is stabilized in the CTD or in the region linking helicase domain and C-terminus of RIG-I. This implies either a conformational change or a protection of the protein by the RNA. AMPPCP does not influence the pattern of the tryptic digest.

Remarkably, a 20-30 kDa fragment of the protein is rapidly (t = 2 min) cleaved off in the tryptic digest digest, resulting in a band of approximately 80 kDa (figure 2.20, blue box). N-terminal sequencing revealed that the 80 kDa band contained a protein fragment starting from amino acid 226. The cleaved portion is completely digested and cannot be found as smaller bands on the gel. This indicates that the tandem CARDs are probably poorly folded and thus very susceptible to protease digestion.

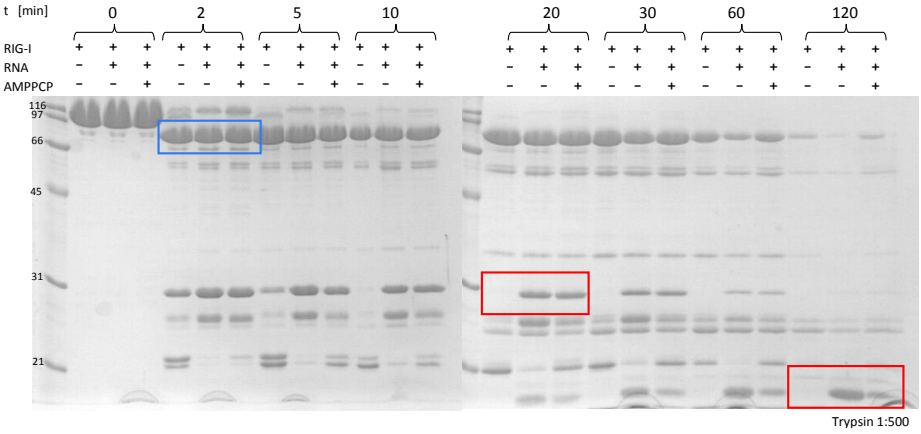


Figure 2.20: *Tryptic digest of RIG-I without RNA, with RNA and with RNA and ATP. Trypsin was added in an 1:500 ratio and samples were taken at the indicated time points. Two fragments (red boxes) are remarkably stable only when RNA is present. CARD domains are very instable as indicated by their rapid cleavage. The approximately 80 kDa fragment (blue box) is starting at residue 226.*

2.1.5 The low resolution structure of human RIG-I derived from small angle scattering

It is common believe that inactive RIG-I has a closed conformation and only upon binding to RNA the protein undergoes a major conformational change to make CARDs accessible for signaling [120, 140, 155]. Furthermore, it is thought that oligomerization of RIG-I is required for signaling [141, 154, 28, 77, 125, 121, 113]. To investigate these hypotheses, Small Angle X-ray Scattering (SAXS) and Small Angle Neutron Scattering (SANS) studies were performed. The experiments were carried out at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at beamline ID14-2 and at the Institute Laue Langevin (ILL), Grenoble, France, at D22, respectively. While SAXS models have a higher resolution resulting from a stronger beam, the contrast variation in SANS allows to identify single components of a complex. Scattering data was collected for different protein concentrations and from the merged curves the radius of gyration (R_g) was determined from the Guinier plot. Then, the maximal distance in the size distribution function was adjusted, so the calculated R_g from the fit would be in agreement with the experimental value. From the fit, 15 to 20 independent models were calculated, each with the parameters R_f <0.01, R_{Los} <0.1 and Dis=0 (measures for model quality). From the generated models, one representative average model was computed.

SAXS

Interpretation of the distance distribution functions The molecular envelopes of full length RIG-I in two different salt concentrations, full length RIG-I with RNA, RIG-IΔCARDS and RIG-IΔCTD were determined with SAXS. Some characteristics of the molecules can already be concluded from the corresponding distance distribution functions (figure 2.21). The histograms represent a count of all distance vectors present in the molecule. Overall, the curves look very similar, indicating similar size and shape of the molecules. The mutant lacking the CTD is notably shorter (maximum distance 110 Å), while the mutant lacking the CARD domains keeps the maximal distance (140 Å) close

to the one observed for the full length molecules (150 Å). This indicates the CTD is at the very tip of the molecule and the are CARDs at the side of the elongated molecule. When shPH RNA is added, the maximal distance stays constant in comparison to RIG-I alone, but more of the long distances are appearing, which indicates the addition of the RNA molecule towards one end of the molecule. No oligomerization can be predicted in presence of shPH RNA, since the maximal distance is constant.

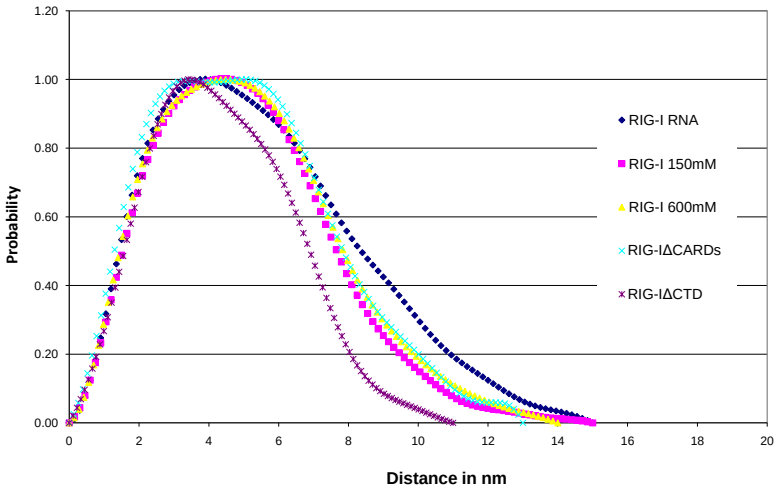


Figure 2.21: *Distance distribution functions of the SAXS and experiments.*

Model interpretation The resulting model of full length RIG-I alone indicates that the protein is an elongated Z-shaped molecule with a length of about 150 Å and a maximal thickness of about 60 Å where the helicase domain is possibly located (figure 2.22). The global shape is independent of salt concentration as shown in the comparison of models from 150 mM NaCl and 600 mM NaCl. To determine the location of the termini in the structure and the placement of CTD and CARDs, we calculated SAXS models of constructs lacking those domains, respectively. From the comparison with the full length model, the CARDs cannot definitively be placed in the envelope. We suspect from the model that they are located at the side of the elongated molecule close to the helicase while the CTD is located at one tip of the molecule, which is in agreement with the obser-

vations based on the distance distribution function. In contrast to previous suggestions, it seems that CARDs and CTD may not touch each other [120].

Modeling of high resolution structures into the SAXS envelope of RIG-I We manually superposed homologous domains of known structures onto the SAXS models, i. e. the helicase of *Pyrococcus Furiosus* hef helicase (PDB:1WP9) and the CTD of RIG-I (PDB:2QFB) (figure 2.23). In the model, the orientation of the CTD was randomly chosen. The hef helicase could be placed in two different ways due to its internal symmetrie. Either it contacts the CTD with the insertion domain or with the first helicase part; we chose the first option, as like this the C-terminus of the helicase is closer to the CTD. The model structures were treated as rigid models, disregarding potential flexibility. The superpositions show that the dimensions of the SAXS models are realistic. Moreover, using the SAXS data of the RIG-IΔCTD construct, the placement of the CTD was unambiguous. Also the global placement of the helicase seems plausible. No distinct CARD domains are distinguished in the full length SAXS envelope, although additional model density can be observed around the helicase. This fact might be due to high flexibility of the CARD domains. It could be speculated that CARDs are unfolded and refold upon a specific event. Such an event could be the binding to an interaction partner or change of physiological conditions. For example, it has been observed that oligomeric state of of the NOD1 CARD is pH dependent [132]. When the RIG-I:shPH RNA complex is measured in SAXS, the model becomes more diffuse due to the flexibility of the RNA, but probably also because of unbound RNA in the solution. A clear location of the RNA cannot be assigned but some extra model density appears close to the CTD of RIG-I. Neither a global domain arrangement nor oligomerization is observed in the presence of RNA.

SANS

To localize the RNA moiety in the complex more precisely, SANS experiments were performed. Due to the different neutron scattering properties of RNA and protein, it is

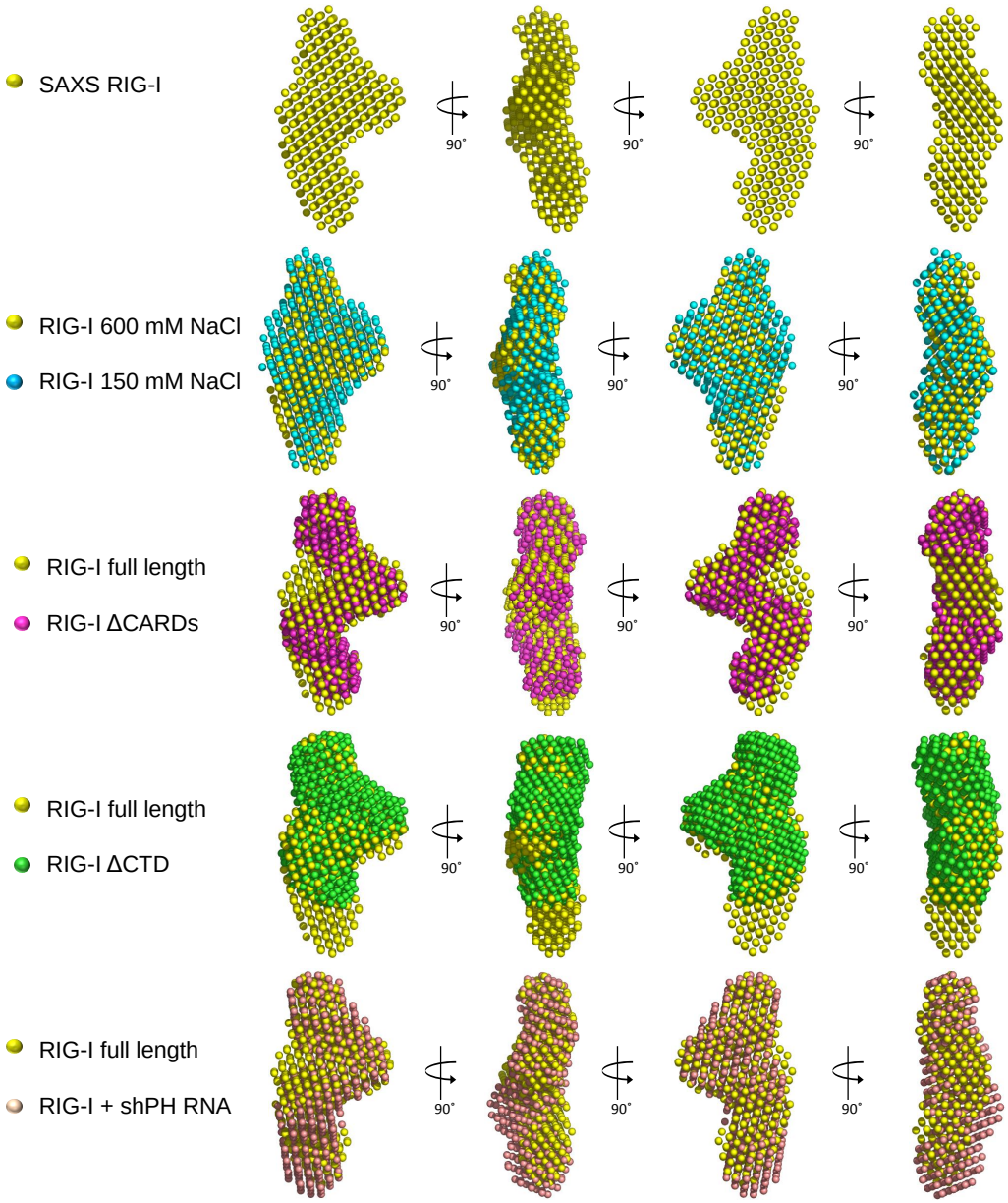


Figure 2.22: *SAXS models of RIG-I.*

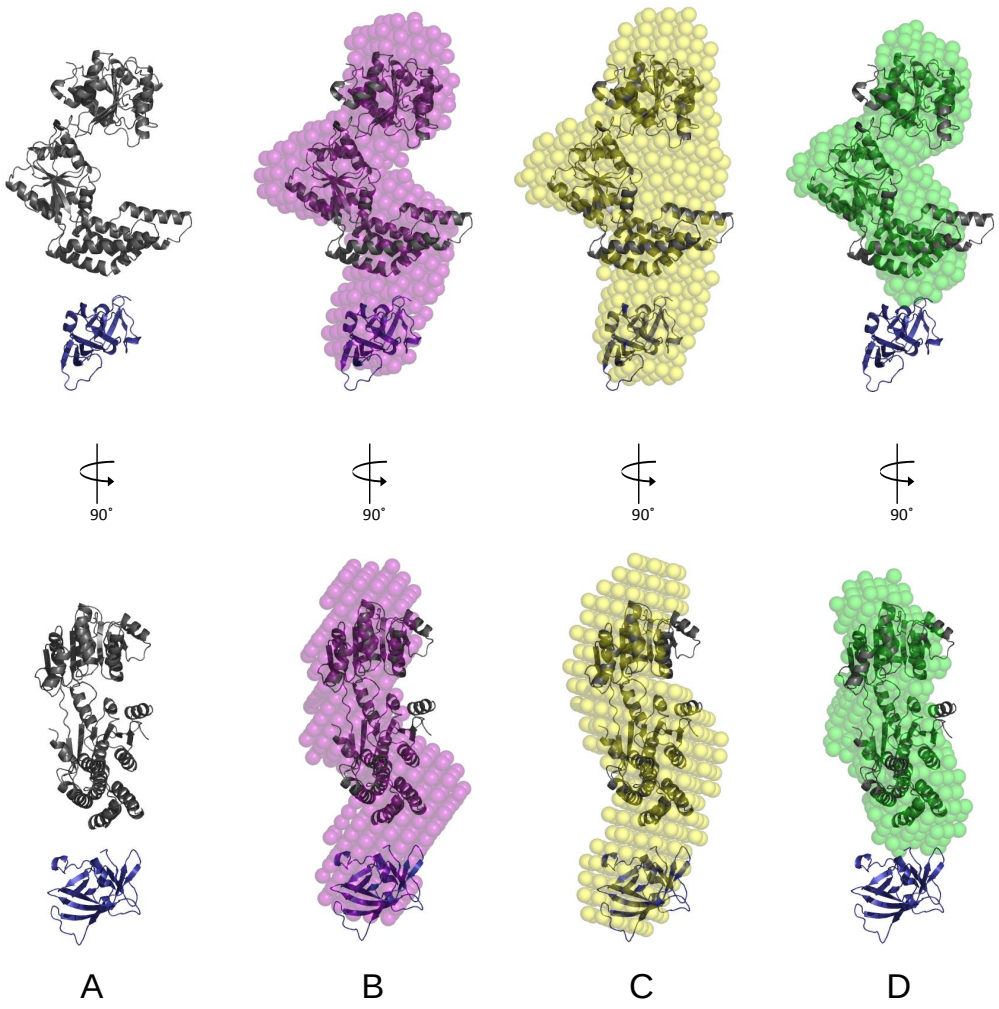


Figure 2.23: *Manual superposition of a RNA helicase domain (1WP9) and the RIG-I CTD (2QFB) crystal structures into the SAXS envelopes. The overall orientation of the domains is random. (**A**) The model structures alone. (**B**) Helicase domain and CTD fit into the SAXS envelope of RIG-IΔCARDs. (**C**) The same domains superposed on the the full length SAXS model. Additional model density can be observed around the helicase domain but distinct CARD domains cannot be discerned. (**D**) An overlay on the SAXS model of RIG-IΔCTD indicates clearly the position of the C-terminal domain.*

possible to choose the H_2O:D_2O ratio of the buffer such that one or other component is "contrast matched out" and does not contribute to the scattering. Contrast matching is achieved at about 43% D_2O for protein and 73% D_2O for RNA. SANS experiments were carried out with full length RIG-I alone in 100% H_2O, 100% D_2O and 73% D_2O buffers, and with RIG-I in complex with RNA in 100% D_2O and 73% H_2O buffers.

Interpretation of the distance distribution functions The distance distribution functions of the SANS experiments are very similar, indicating similar shape for all envelopes (figure 2.24). The 100% D_2O curve of the complex indicates a larger maximal distance. Also more large distances are present in comparison with the 100% D_2O curve of the protein alone pointing to the addition of RNA at one tip of the elongated molecule. Remarkingly, curves of RIG-I alone in 100% D_2O and of the complex in 73% D_2O (RNA matched out) are strikingly similar, an indicate that the overall shape of the protein molecule is not changing after binding of shPH RNA. The 100% H_2O data was collected to ensure that the buffer exchange to D_2O did not induce artifacts. Since the distance distribution function of the 100% H_2O and 100% D_2O data are essentially the same, we can suppose that the deuterium buffer did not harm the protein. However in SANS, measurements in H_2O buffer are more noisy due to enhanced background signal, which leads models with less details.

Model interpretation Comparing first the 100% H_2O SANS model with the SAXS model, we see a more compact envelope, due to noise in the SANS raw data (figure 2.25). Less density in the C-terminal tip of the 100% H_2O SANS model compared to the SAXS model might indicate flexibility of this part of the protein. The SANS model of the protein alone in 100% D_2O is essentially the same as the SAXS model. The model of RIG-I with shPH RNA measured in 73% of D_2O adopts very similar conformation as the previous ones. Although RNA is physically present, only the protein moiety is visible under this condition. As the 73% D_2O model with RNA (complex but only protein visible) and the 100% D_2O (monomer without RNA) resemble, we conclude that no global domain

rearrangement is taking place in the molecule upon RNA binding. Again, there is less density visible for the C-terminal tip, due to higher noise in this model. This indicates flexibility of this part of the protein. The 100% D_2O model of the complex is very noisy probably because of the free RNA in the solution, but reflects the overall expected shape.

Conclusions from SAXS and SANS structures

From SAXS and SANS data, it can be concluded that RIG-I does not undergo a global domain rearrangement or multimerization after binding to the shPH RNA. The RNA binding site cannot clearly be localized but seems to be towards the tip of the elongated molecule. Likewise, the placement of the CARDs is not clear from the model, probably due to high flexibility. The CARDs might have special folding properties depending on pH or other additional factors or interaction partners. It would be interesting to perform additional SAXS measurements with the full length protein at different pHs. The CTD and helicase domain could be located in the low resolution structure.

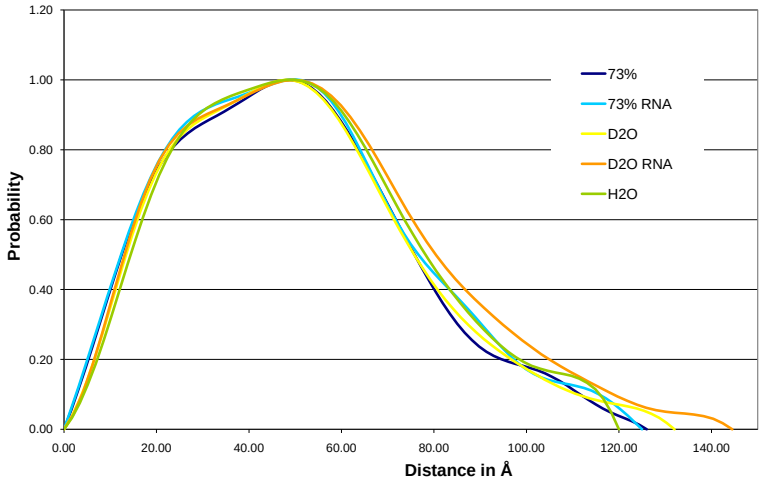


Figure 2.24: *Distance distribution functions of the SANS experiments.*

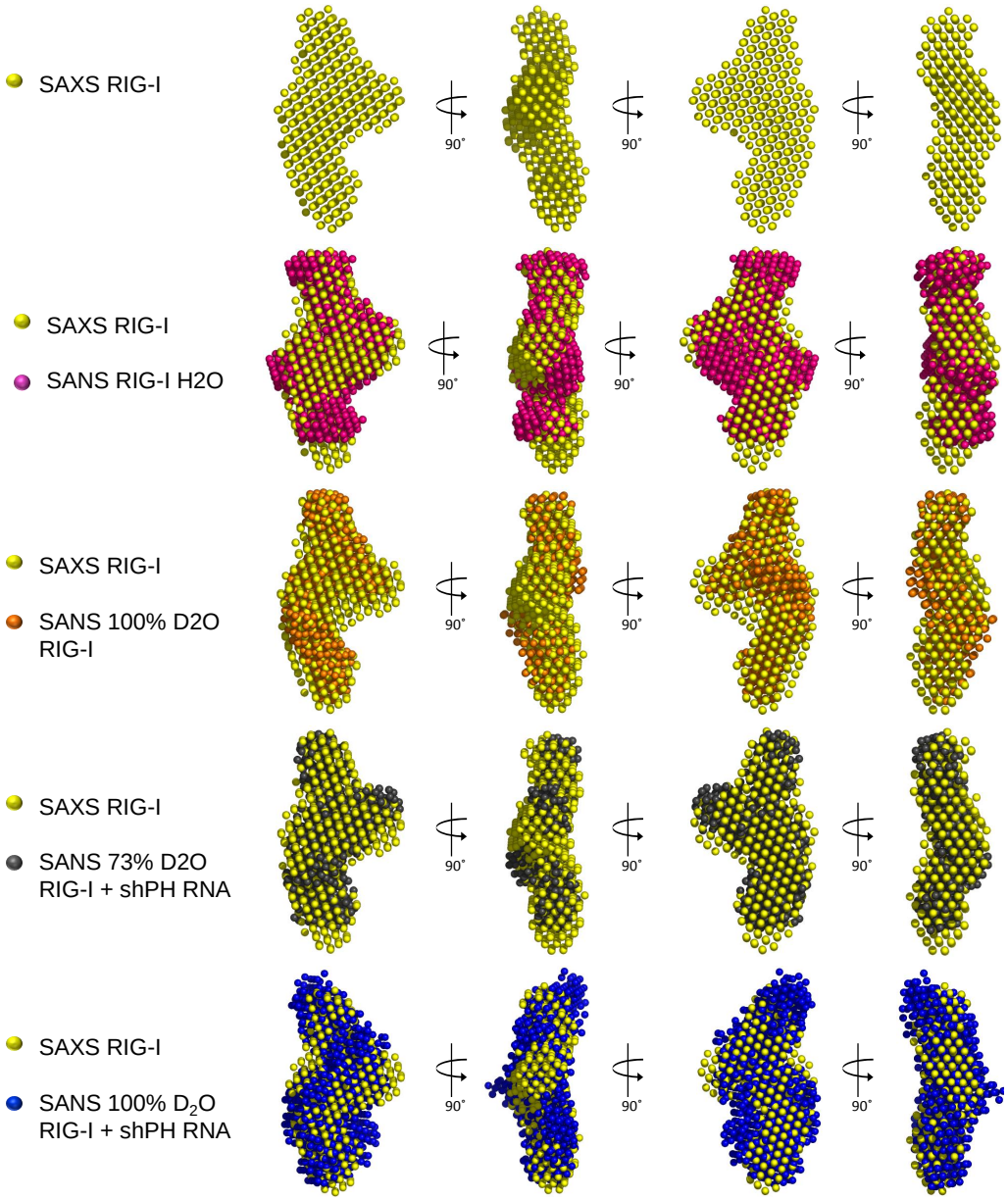


Figure 2.25: *SANS models of RIG-I.*

2.1.6 RIG-I and ubiquitin

In contrast to the proposed model of multimerization of RIG-I upon RNA binding [141, 154, 28, 77, 125, 120, 113, 157], which will be discussed in chapter 3 Discussion, we could not observe dimers or oligomers of RNA bound RIG-I. However, it is possible that multimerization of RIG-I occurs rather at a later step of activation then immediately upon RNA binding. Different studies claim direct ubiquitination of RIG-I by TRIM25 necessary for signaling [44, 46] or the interaction of RIG-I with K63-linked poly-ubiquitin chains generated by TRIM25 [157]. Thus, RIG-I multimerization could happen at the step when the CARDS interact with poly-ubiquitin in the RIG-I:RNA:ubiquitin complex or more downstream on adapter level, in the RIG-I:RNA:ubiquitin:MAVS complex. For further investigation, we tried to achieve covalent RIG-I ubiquitination or to establish a RIG-I:K63-linked-tetra-ubiquitin complex *in vitro*.

In vitro ubiquitination of RIG-I

Gack and co-workers demonstrated that RIG-I CARDs are ubiquitinated by the E3 protein ligase TRIM25 [44]. They could also show *in vitro* ubiquitination of RIG-ICARDs. We tried to reproduce the published experiment, incubating E1 ligase (UBA1), E2 ligase (UbcH5a) and purified TRIM25 with purified CARDs and with full length RIG-I in presence and absence of RNA (shPH and MVL). The His-tagged E1, E2 and ubiquitin were purchased and TRIM 25, RIG-I and the RNAs produced recombinantely. The ubiquitination reactions were analyzed on an SDS gel, but poly-ubiquitination pattern (protein + n8 kDa) of RIG-I or CARDs could not be detected (figure 2.26). However, in the reactions containing all components with and without RNA, high molecular weight complexes which do not enter the gel, can be observed (black arrows).

To identify the components present in those high molecular weight bands, western blot analysis was performed with antibodies against the His-tag and against ubiquitin (2.27). The high molecular weight bands contain both, His-tagged species as well as ubiq-

uitin. Mass spectroscopy analysis identified ubiquitin and TRIM25 present in these bands. This can be due to the temporary covalent linkage of the E3 ubiquitin ligase TRIM25 to ubiquitin. Moreover TRIM proteins are known to form high oligomeric structures via their coiled coiled domains [98] and TRIM25 can appear *in vivo* as diffuse aggregates in the cytoplasm [115].

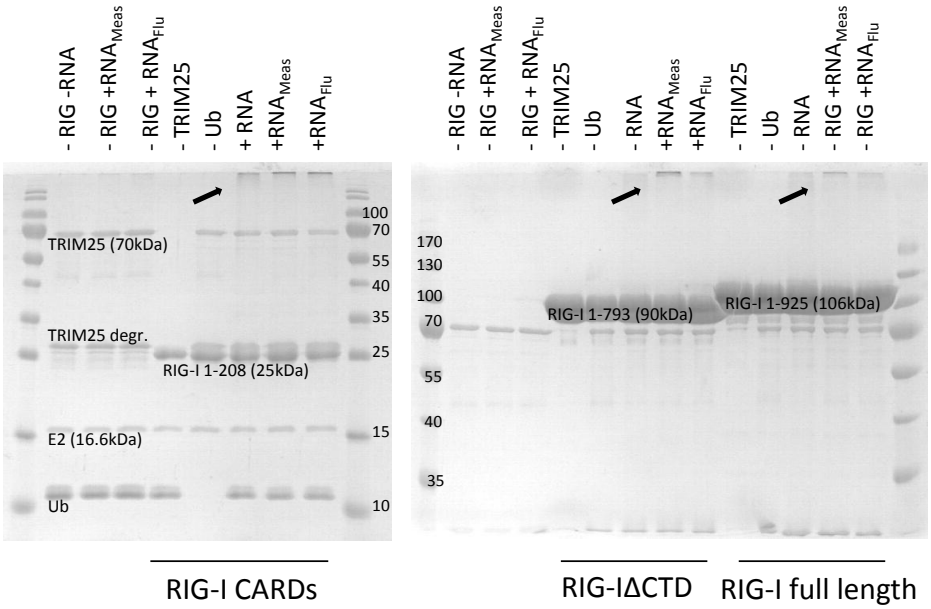


Figure 2.26: *An in vitro ubiquitination experiment was performed on three different constructs of RIG-I (CARDs, RIG-IΔCTD and full length RIG-I). The reactions were analyzed on SDS PAGE.*

RIG-I interaction with free polyubiquitin chains

In contrast to Gack and co-workers, Zeng and co-workers have shown that the interaction of RIG-I with unanchored polyubiquitin chains is required for RIG-I activation [157]. In particular, the CARDs of RIG-I bind K63-linked polyubiquitin chains of three or more ubiquitin units. To analyze weather RIG-I can oligomerize in presence of the ubiquitin signal, we wanted to establish the interaction *in vitro*.

Structurally, K63-linked chains are very similar to linear head-to-tail polyubiquitin chains [72]. In a first approach, co-expression and co-purification of RIG-I CARDs with

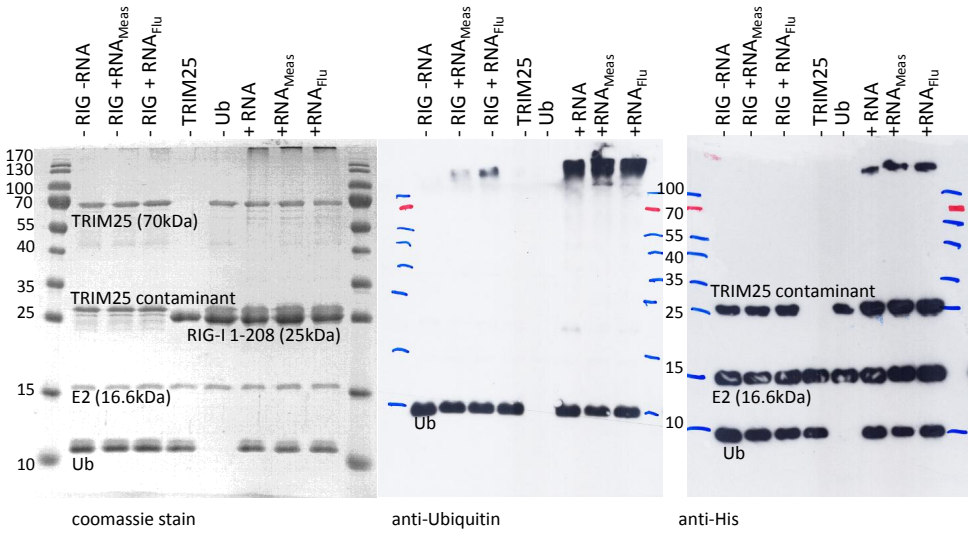


Figure 2.27: *The in vitro ubiquitination experiment on RIG-I CARDs was analyzed on SDS PAGE and on anti-ubiquitin and anti-his western blots.*

linear polyubiquitin chains (bi-, tri- and tetra-polyubiquin) was tested and secondly the interaction of RIG-I and RIG-I CARDs with commercially available K63-linked polyubiquitin chains was analyzed by SEC-MALLS experiments.

Interaction with linear polyubiquitin chains To analyze the interaction of RIG-I with linear polyubiquitin chains, we co-expressed and co-purified those with the CARDs of RIG-I. Clones for di-, tri- and tetra-ubiquitin chains (Ub_2, Ub_3 and Ub_4) were provided by David Komander, Cambridge. His-tagged RIG-I 1-226 was co-expressed with the ubiquitin chains and the cell lysate was passed through a nickel affinity column, but untagged linear polyubiquitin chains did not co-purify with his-tagged RIG-I, suggesting they do not interact. (figure 2.28, the result for Ub_2 is not shown).

Interaction with K63 linked polyubiquitin chains In a second approach, commercially available enzymatically linked K63-linked tetra-ubiquitin ($K63Ub_4$, Boston Biochem) was incubated with RIG-I CARDs and with full length RIG-I:RNA complex. The mixture was analyzed by MALLS (figures 2.29 and 2.30). In the experiment, the protein:RNA complex has an apparent mass of 117 kDa and no further shift can be observed

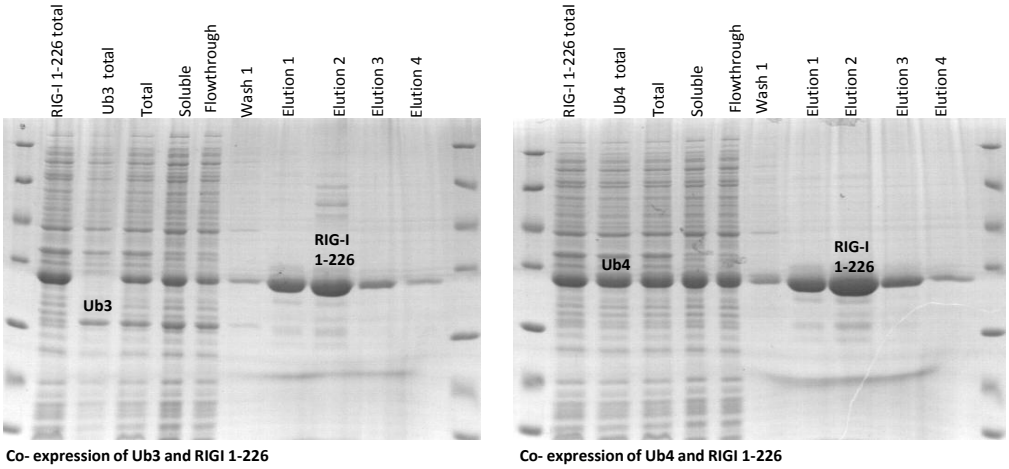


Figure 2.28: *Untagged linear ubiquitin chains (Ub_3: 3 ubiquitin units, Ub_4: 4 ubiquitin units) were co-expressed with the his-tagged double CARD of RIG-I and purified on a nickel affinity column. The lysate is compared to lysate only expressing the ubiquitin chain and CARDs, respectively. No interaction of the linear ubiquitin chains with RIG-I CARDs can be observed.*

in the presence of $K63Ub_4$. Equally, the expected mass shift (to at least 60 kDa for a 1:1 complex) cannot be observed when $K63Ub_4$ is added to the CARDs of RIG-I (1-226). The lack of interaction might be due to the salt concentration (150 mM NaCl) which might be too elevated to observe such a transient signaling complex. Alternatively, further factors might be necessary for the interaction.

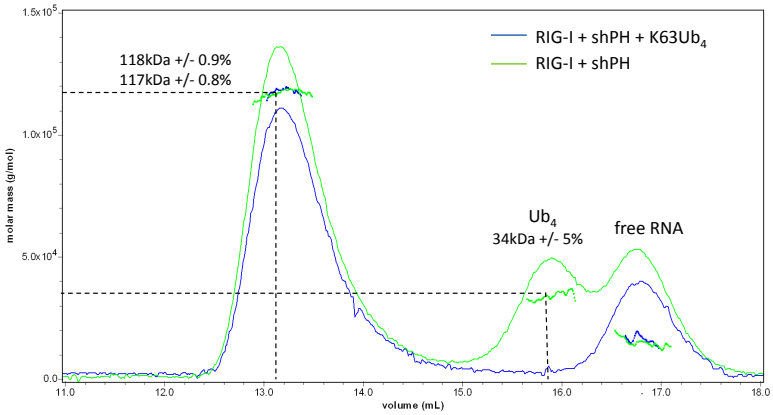


Figure 2.29: *Full length RIG-I does not interact with $K63Ub_4$ chains in presence of RNA (SEC-MALLS data)*

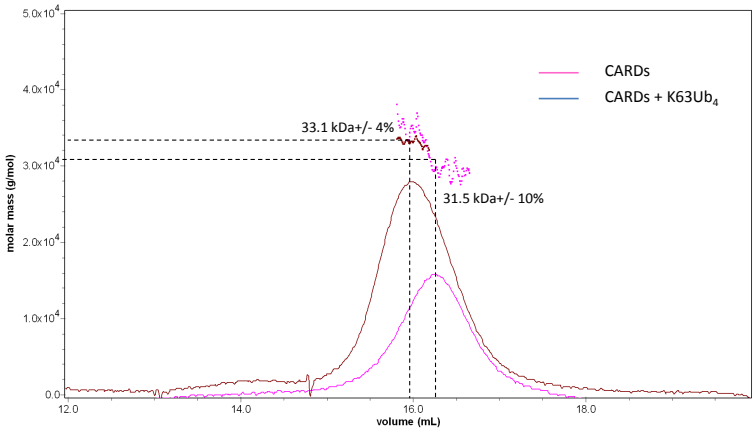


Figure 2.30: *RIG-I CARDs (1-226) do not show interaction with K63 linked tetra ubiquitin in MALLS. Since sizes of CARDs and K63Ub$_4$ are very similar, they do not appear as single peaks but as one large peak.*

Conclusions of RIG-I-ubiquitin assays

In our hands, RIG-I did neither interact with poly-ubiquitin chains (linear or K63 linked) nor was it possible to ubiquitinate RIG-I *in vitro*. The interaction with polyubiquitin chains might be very transient and weak, so no stable complex is formed in the salt and buffer conditions used. Additionally, it is possible that co-factors (ATP, ADP, Mg^{2+}) are needed which are not present in the *in vitro* experiment. The reproduction of the *in vitro* ubiquitination assay from Gack *et al.* [44] with RIG-I CARDs was not successful. This might be due to failure of one of the ubiquitination enzymes. Probably the quality of the purified TRIM25 was not sufficient. A positive control for this experiment is difficult to establish since the E3 ligase is very specific for the substrate and not commercially available. High molecular weight oligomers of TRIM25 bound to ubiquitin were found in the reactions.

2.1.7 Preliminary crystallographic analysis of RIG-I from mallard duck

The choice of the organism "mallard duck"

Since human RIG-I was failed to crystallize, we decided to express, purify and crystallize RIG-I from another species. Several constructs of platypus (*Ornithorhynchus anatinus*), an ancient mammal with a very conserved genome, were expressed in insect cells but the resulting protein was not soluble (data not shown). As a further species, RIG-I from mallard duck (*Anas platyrhynchos*) was tested. Birds can be infected with all strains of the influenza virus. It was observed that ducks are often resistant to influenza viruses capable of killing chickens. One reason for the lacking immunity of chicken against influenza could be that RIG-I, the sensor for the influenza viral RNA, is absent in chicken but present in ducks. The DF-1 chicken cell line is naturally deficient in RIG-I. When transfected with *in vitro* transcribed 21mer 5'-PPP RNA, a potential RIG-I ligand, it does not show any immune response while transfection with polyI:C induces interferon via MDA5, as this receptor is present. When the DF-1 cells were additionally transfected with duck RIG-I the immune response to the 5'-PPP RNA could be reconstituted [8].

Mallard duck RIG-I (dRIG-I) has 53% amino acid homology with the human homologue and consists of 933 amino acids. Domain prediction reveals the expected tandem N-terminal tandem CARD domains and a helicase domain, consistent with the mammalian structure. dRIG-I is predicted to be a ligand-dependent ATPase, and the Walker A ATP-binding motif is conserved. An alignment of human and mallard duck RIG-I is shown in figure 2.31. Within the C-terminal region, the hydrophobic core and the four lysine residues implicated in ligand binding in human RIG-I, K858, 861, 888 and 907 are conserved. However, in the CARDs, residues T55 and K172, critical for interaction and polyubiquitination by TRIM25 and activation of MAVS are not conserved in mallard duck. This key feature lead to the idea to choose dRIG-I for crystallization, since regulation of human RIG-I by polyubiquitination is a hardly controllable factor for biochemical exploration and crystallograpy.

Figure 2.31: *Alignment of human and mallard duck RIG-I protein sequence. The tandem CARDs, helicase 1st, insertion and 2nd part and CTD are annotated. Zn-binding residues in CTD are marked with a yellow star, the residues responsible for fixing the ligand phosphate are marked with a green star. Furthermore interesting residues are annotated (human residue numbers): S8 a phosphorylation site; T55 responsible for interaction with TRIM25; K127 to which the ubiquitin is linked; K270 necessary for ATPase function; Q299 a residue responsible for RNA binding in the helicase.*

Expression, crystallization and preliminary analysis

The gene of mallard duck RIG-I, codon optimized for the expression in *Trichoplusia ni* insect cells, was cloned into pFastBac Htb and expressed in baculovirus expression. The expression yield was approximately 10 mg per 1 L of insect cell culture. After a standard purification procedure (first nickel affinity column, TEV cleavage, second nickel affinity column, size exclusion chromatography), the protein eluted homogeneous as a monomer (figure 2.32 A and B). Purified full length dRIG-I interacts with shPH RNA in a similar fashion to human RIG-I, as analyzed by SEC-MALLS (figure 2.33).

Crystallization trials were set up with 10 to 20 mg/ml of protein, containing 5 mM tris(2-carboxyethyl)phosphine (TCEP) in presence and absence of shPH RNA. Mallard duck RIG-I alone crystallized in two conditions: (i) 0.1 M KCl, 0.01 M Mg_2Cl, 0.05 M Tris pH 8.5, 30%v/v polyethylene glycol (PEG) 400 or (ii) 20% w/v polyethylene glycol 3350 and 0.2 M potassium nitrate, sodium nitrate, lithium nitrate or potassium iodide, respectively (figure 2.34 B and C). However, crystals were difficult to reproduce, grew in skin and stuck to the plastic or glass surface of the crystallization device. Crystals grown in sitting drops had a superiour quality compared to hanging drop crystals. They diffracted to a resolution of 3.6-4 Å. A native dataset and a dataset with a wavelength of the absorption peak for zinc were collected on the same crystal. The crystals are of space group $P4_12_12$ (92) or $P4_32_12$ (96) with cell dimensions of 164.36, 164.36, 153.31 Å. The data statistics are shown in table 2.1. The cell dimensions correspond to the longest distance observed in the SAXS and SANS measurements for a monomer of human RIG-I (approximately 150 Å, see section 2.1.5). The X-ray absorption spectrum (figure 2.34) indicates the presence of zinc in the crystals consistent with the fact that the CTD of RIG-I contains one Zn^{2+} atom.

The structure could not yet been solved by molecular replacement although various models of homologous helicase domains, parts of helicase domains, CARD domains and

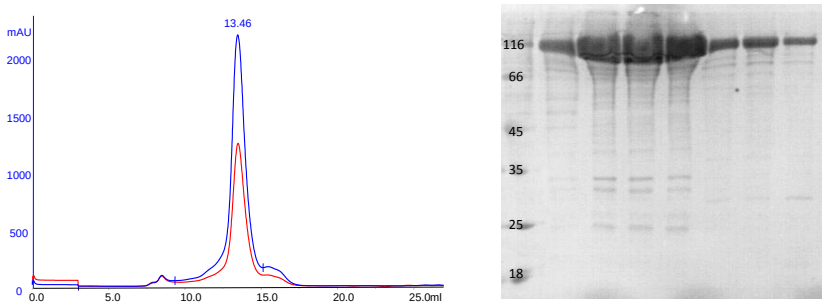


Figure 2.32: *Purification of dRIG-I. (**A**) Size exclusion chromatography profile. (**B**) SDS-PAGE of the peak fractions.*

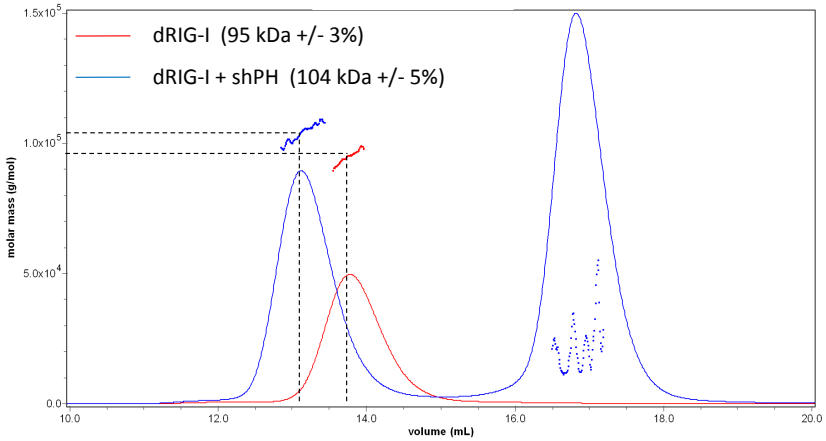


Figure 2.33: *dRIG-I interacts with shPH RNA as observed by MALLS.*

Table 2.1: *Data collection and refinement statistics.*

Data collection	
Beamline	ESRF,ID23EH1
Detector	ADSC Q315R
Wavelength (Å)	1.27985
Space group	$P4_12_12$ (92) or $P4_32_12$ (96)
Cell dimensions (Å)	164.36, 164.36, 153.31
Cell angles(°)	90, 90, 90
Resolution (Å)	4.0-50 (4.0-4.14)
Completeness (%)	99.8 (99.9)
R_{merge} (%)	9.9 (69.4)
I/σ (I)	10.66 (2.48)

[1] The information in parentheses refers to the highest resolution shell

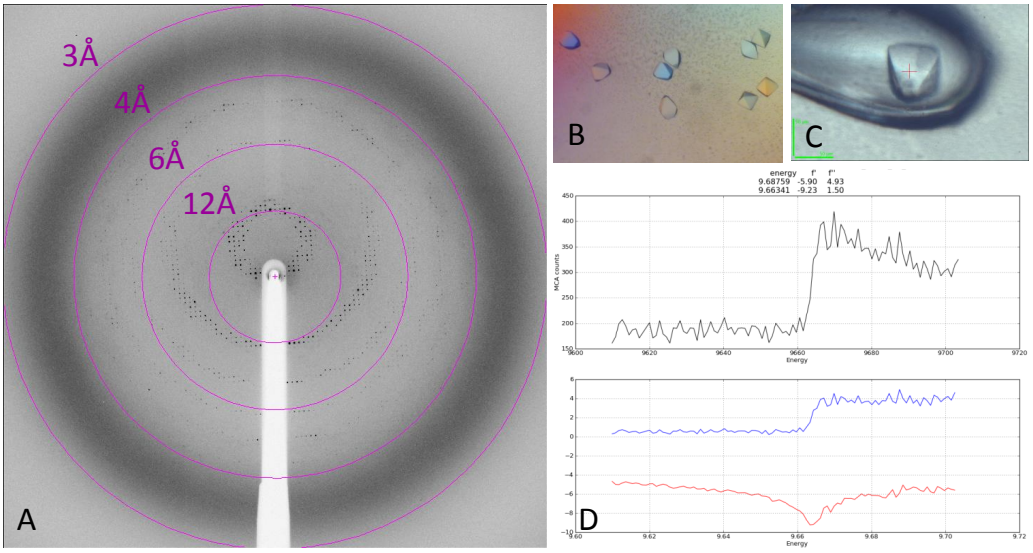


Figure 2.34: *Preliminary results of crystallization of RIG-I from Anas platyrhynchos (**A**) Diffraction pattern. (**B**) Crystal drop. (**C**) Mounted crystal. (**D**) Absorption spectrum for Zn.*

the RIG-I CTD were tested as models. Trials for expression with seleno-methionine (Se-Met) replacement and heavy atom soaks of native crystals are currently being carried out. Se-Met incorporation in the baculovirus expressed protein was achieved following a protocol of Cronin and co-workers [27]. For this, insect cells are infected with the virus as usual. After a certain time period the medium is exchanged to methionine free medium. After further incubation to deplete residual methionine, seleno-methionine is added and expression occurs. The incubation times and amount of Se-Met have to be individually optimized for different proteins. The incorporation of Se-Met can be analyzed by mass spectroscopy of the purified protein. We grew crystals of a batch of protein for which the medium was changed after 24 h to medium containing 100 mg/L Se-Met. However mass spectrometry results indicated two species, a fraction of native RIG-I and a fraction of protein with 17 of 25 possible Se-Met present in average. The combination of incomplete incorporation, presence of native protein, small crystals and a low resolution of approximately 8 Å did not yet allow to obtain phase information. A crystal and its X-ray absorption spectrum are shown in figure 2.36 . A second batch with 18 h infection time prior to medium exchange, 4 h incubation time without any methionine and subsequent

addition of 150 mg/L Se-Met resulted in better Se-Met incorporation (23 of 25 possible in average) but still there is native protein present (figure 2.35 A). The native fraction contains most probably protein expressed before the medium exchange, even if protein production cannot be observed on SDS-PAGE after 24 h post-infection (figure 2.35 B). We are optimizing now the time point of medium exchange to deplete native protein expression. We will also attempt to obtain larger crystals by seeding, which might be critical for successful structure determination.

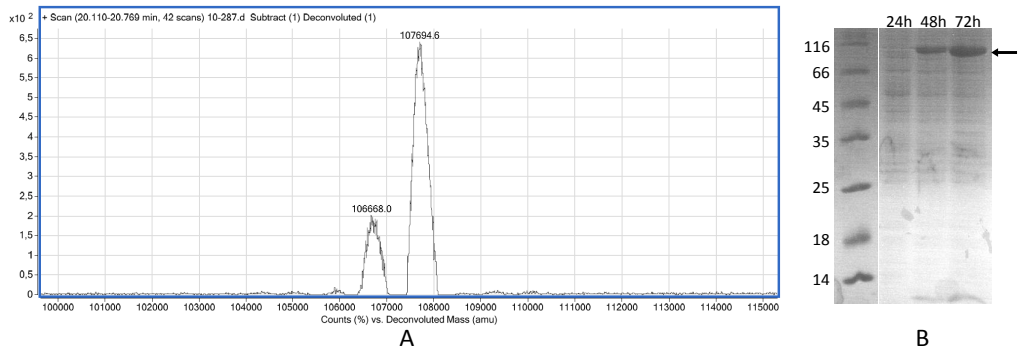


Figure 2.35: *Seleno-methionine (Se-Met) incorporation into dRIG-I. (**A**) Mass spectrometry detects two species (Liquid chromatography time of flight (LC-TOF)). (**B**) Expression profile of dRIG-I for a Se-Met incorporation protocol.*

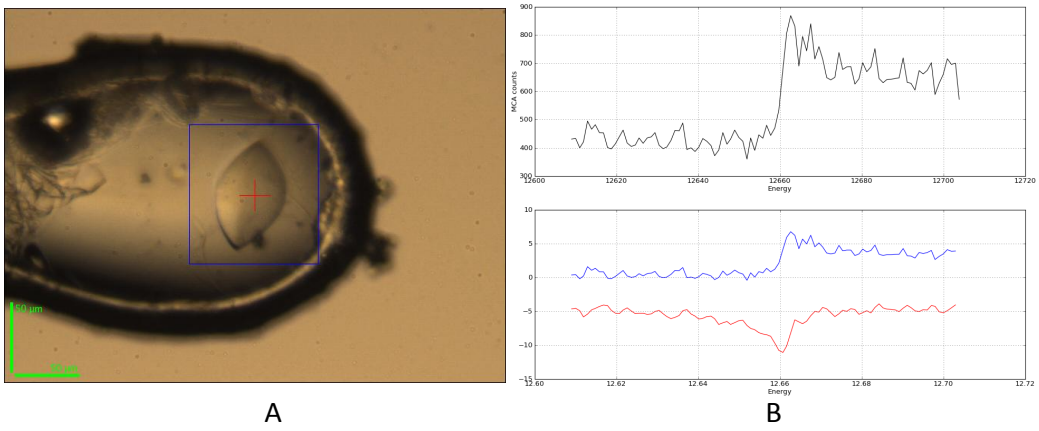


Figure 2.36: *(**A**) Crystal of seleno-methionine dRIG-I. (**B**) X-ray absorption spectrum of the same crystal*

2.2 The structure of the TRIM25 PRYSPRY domain

2.2.1 Protein expression

The full length E3 ubiquitinase TRIM25 could never be expressed in a soluble form neither in *E.coli* nor in insect cells. A MBP-tagged construct of full length TRIM25 can be expressed in *E.coli*. But after purification, the protein elutes as soluble aggregates in the void volume of an S200 size exclusion column and a contaminant is co-eluted with TRIM25. Different constructs of the PRYSPRY domain of TRIM25 were designed since it has been proposed to be the interacting domain with the RIG-I CARDs [44].

2.2.2 Structure solution

Initially, TRIM25 PRYSRPY domain was expressed to characterize its interaction with RIG-I CARDs and to co-crystallize the complex. However, the results from surface plasmon resonance (SPR) measurements are difficult to interpret (see section 2.2.5) and complex crystals could not be obtained yet. Nevertheless, a fragment of TRIM25 PRYSPRY comprising residues 407-630 alone crystallized. The initial crystals diffracted poorly. The unit cell constants predicted a large number of molecules per asymmetric unit and a high solvent content. Thus, shorter constructs were designed. A fragment containing residues 434-630 could be expressed and purified to homogeneity (figure 2.37). It was crystallized by vapor diffusion in sitting drops mixing 15 mg/ml protein with reservoir solution in an 1:1 ratio. The reservoir solution contained 0.1 M bis-tris pH 6.5 and 2 M ammonium sulfate. To phase the structure, a crystal grown under native conditions was transferred in the same solution containing 2 mM ethylmercuric chloride (EMC) and soaked for 8 h before freezing. The structure could be solved by SIRAS (single isomorphous replacement with anomalous scattering) at a resolution of 2.3 Å using one native and one mercury soaked crystal. Six Hg sites were found, refined with SHARP and used for phasing. The map was improved by 3-fold non-crystallographic averaging and an initial model was built by ARP/WARP [32, 73].

The native crystals belong to the $P32_1$ space group with cell dimensions of 152.07, 152.07 and 68.04 Å. There are three molecules of TRIM25 PRYSPRY domain in the asymmetric unit.

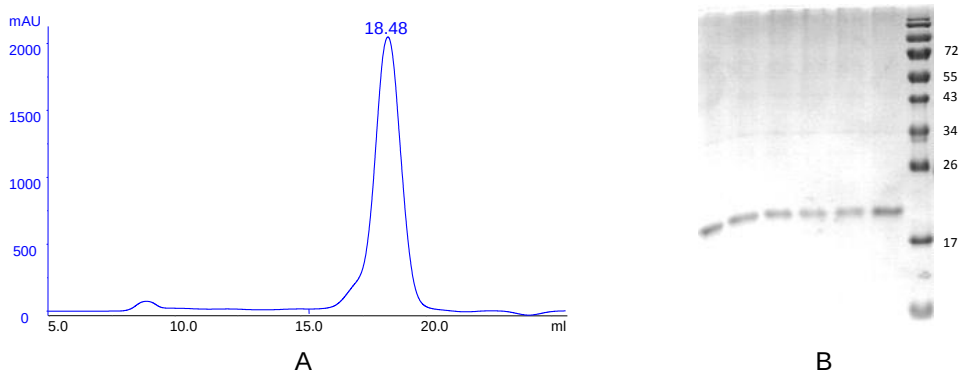


Figure 2.37: *Purification of TRIM25 434-630. (**A**) Profile of size exclusion chromatography and (**B**) corresponding peak fractions on SDS PAGE.*

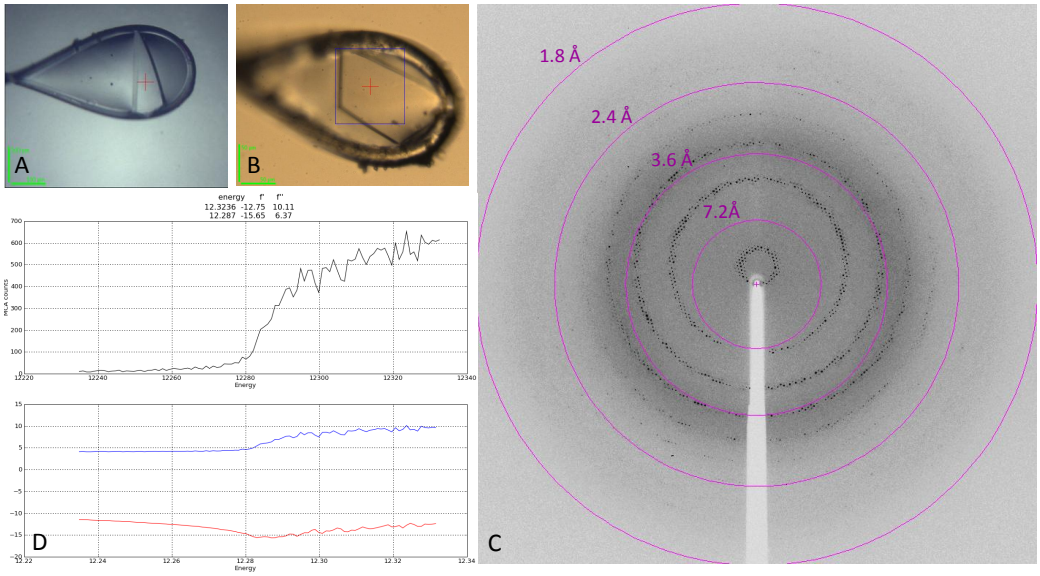


Figure 2.38: *Structure solution of TRIM25 PRYSPRY.(**A**) Native crystal. (**B**) Crystal soaked with ethylmercuric chloride (EMC) which was used for structure solution (**C**) Diffraction pattern of the native crystal (**D**) Energy absorption spectrum of the EMC soaked crystal.*

Table 2.2: *Data collection and refinement statistics.*

Data collection	native	Hg derivative
Beamline (ESRF)	ID14-4	ID14-4
Detector	ADSC Quantum Q315r	ADSC Quantum Q315r
Wavelength (Å)	0.91850	1.00000
Space group	*P*321 (150)	*P*321 (150)
Cell dimensions (Å)	152.07, 152.07, 68.04	152.31, 152.31, 68.74
Cell angles (°)	90, 90, 120	90, 90, 120
Resolution (Å)	2.30-100 (2.30-2.38)[1]	3.0-50 (2.50-2.59)
Completeness (%)	99.5 (95.6)	98.6 (98.8)
R_{merge} (%)	13.7 (88.6)	9.8 (33.7)
I/σ (I)	12.54 (2.17)	12.14 (4.83)
Refinement		
Resolution (Å)	2.30-50 (2.30-2.36)	
Total reflections in working set	38324 (1954)	
Total reflections in test set	2041 (96)	
No. of non-hydrogen atoms	5360	
No. of solvent waters	610	
Ligands	7 SO_4	
R_{work}	0.163 (0.216)	
R_{free} (%)	0.204 (0.298)	
Mean B factor, protein atoms (Å^2)	20.424	
r.m.s.d. from ideal bond length (Å)	0.016	
r.m.s.d. from ideal bond angel (°)	1.551	
Ramachandran plot favored[2] (%)	97.4	
Ramachandran plot outliers (%)	0.0	

[1] The information in parentheses refers to the highest resolution shell

[2] As calculated by MOLPROBITY [18]

2.2.3 The overall fold

The structure comprises a globular fold with two perpendicular N-terminal α-helices (α1 and α2) followed by the PRYSPRY domain. The PRYSRPY domain forms a jellyroll β-sandwich which is the hydrophobic core of the structure. An overview on the structure and a topology diagram are shown in figure 2.39. The β-sheets contains 13 β-strands of which the first three belong to the PRY domain and 11 are part of the SPRY domain. β-strands 3, 12, 5, 6, 7 and 8 form one sheet (sheet A) and strands 1, 2, 13, 4, 9, 10 and 11 build up to the second sheet (sheet B). The helices α1 and α2 pack on the outside of the β-sandwich to sheet B. In homologous structures, the surface of sheet A, together with some loop regions serves as the interaction surface with various binding partners. Those variable loops (annotated VL1-6) are analogous to the six "complementary determining regions" (CDR-loops) in antibodies and the overall fold of the PRYSPRY domain resembles an extended immunoglobulin fold.

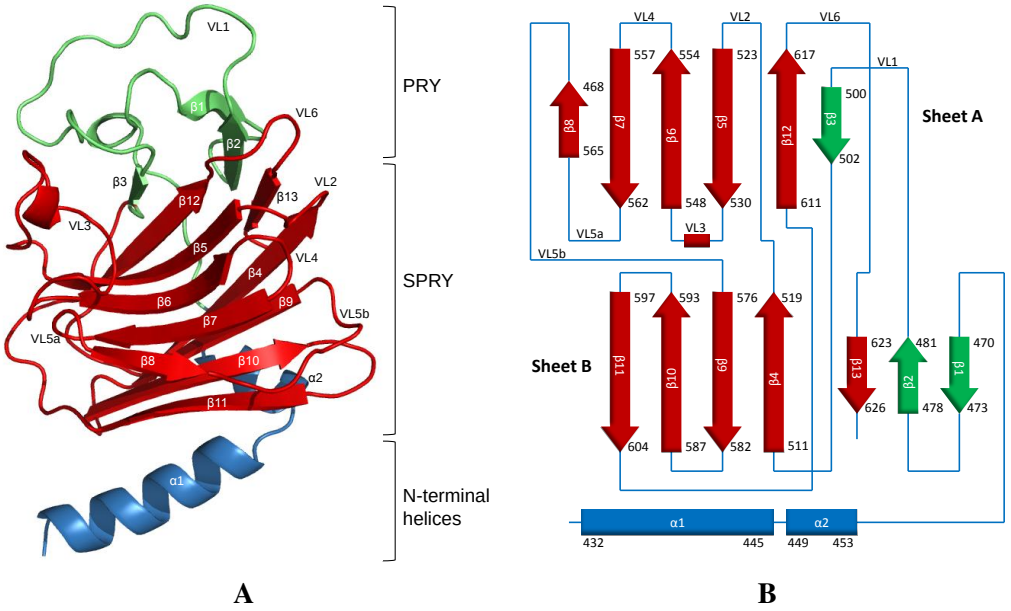


Figure 2.39: *The structure of TRIM25 PRYSPRY domain. (**A**) Cartoon model of the overall fold. The N-terminal helices (blue) are followed by the PRYSPRY domain (PRY in green, SPRY in red) (**B**) Topology diagram of the solved structure. Variable loop regions (VL) are indicated.*

2.2.4 Putative interaction site

PRYSPRY domains are a platform various protein-protein interactions. To identify putative interaction surfaces on the TRIM25 PRYSPRY domain, two strategies were followed: the comparison with homologous structures and the search for conserved surface residues. A search with the DALI server [57] found murine and human TRIM21 (2VOK, 2VOL, 2IWG), the SPRY domain of pyrin (2FBE, 2WL1) as similar structures (rmsd < 2 Å). Both TRIM21s are well characterized in complex with an antibody F_c fragment [70, 61]. In mouse TRIM21, the binding pocket has been analyzed by mutagenesis studies.

TRIM21 and TRIM25 PRYSPRY domains were structurally aligned and the TRIM21:F_c interaction residues were mapped on the surface of TRIM21 (figure 2.40 A, interaction residues in red). Mapping the same residues on the surface of the TRIM25 PRYSRPY domain shows a different distribution as these residues are not conserved and orientated differently, although the protein backbone is well aligned (figure 2.40 B). Futhermore, TRIM21 possesses a deep binding pocket whereas this feature is much less pronounced in TRIM25.

In a second approach the surface conservation of TRIM25 PRYSPRY was analyzed. For this a protein sequence alignment of the TRIM25 PRYSPRY domains of nine different species was generated and the percentage of conservation per residue was mapped on the surface of the human structure (figure 2.40 C). In contrast to TRIM21, not the residues surrounding the pocket but the base of the pocket are conserved. The mode of interaction of TRIM25 PRYSPRY to a binding partner presumably differs from the TRIM21:F_c complex. Weinert and co-workers discussed the variability of the binding pocket in the PRYSPRY domains of sRFPL1, TRIM21 and pyrin [148]. TRIM21 recognizes a hairpin loop whereas rSRPL1 is unlikely to be able to interact with a hairpin like structure (it crystallized with the C-terminal tail of a symmetry related molecule in the central cavity). Weinert postulated that the putative interaction partner for pyrin cannot resemble

the peptides bound to sRFPL1 and TRIM21. The structure of *Drosophila* GUSTAVUS has a less pronounced central cavity. It is a protein interaction module binding VASA with its immunoglobulin like surface. GUSTAVUS interacts also also with the elongin BC complex using a different surface of the molecule, corresponding to the area where the N-terminal helices are located in TRIM25 [150].

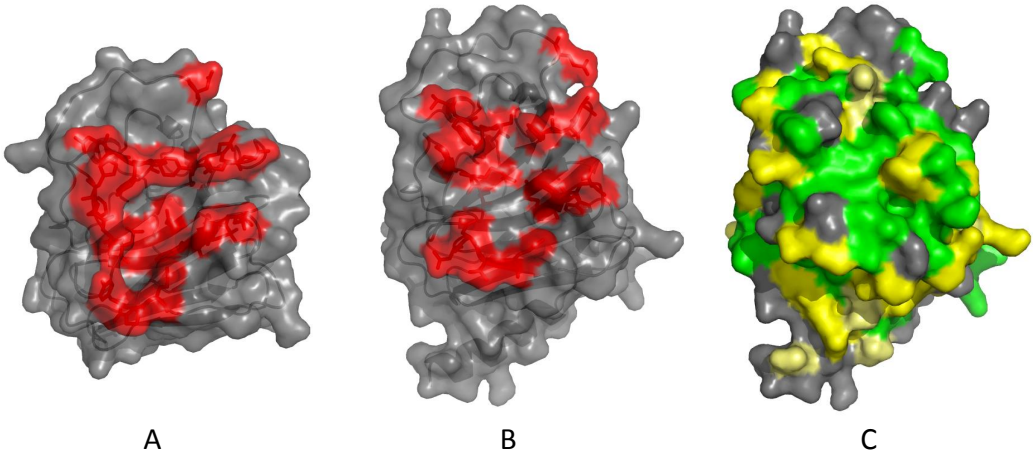


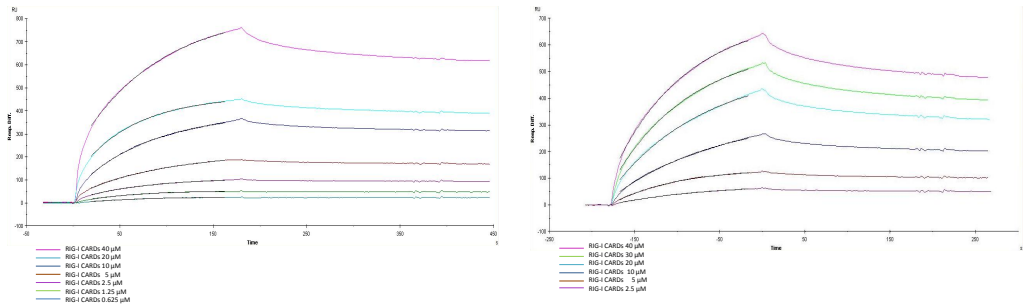
Figure 2.40: *(**A**) Surface representation of murine TRIM21 PRYSPRY. Residues involved in F_c-binding are colored in red. (**B**) Surface representation of TRIM25 PRYSPRY. The equivalent residues from TRIM21 are mapped in red on the surface of TRIM25. (**C**) Surface conservation of TRIM25 PRYSPRY. From an alignment of nine species, conservation was calculated and mapped on the surface (paleyellow >50%, yellow >70%, green = 100% conservation.)*

2.2.5 Interaction of TRIM25 PRYSPRY with RIG-I CARD

It has been shown that the interaction of TRIM25 with RIG-Iis mediated by respectively the PRYSRPY domain and CARDs [44]. A key residue for the interaction is T55 in the first CARD of RIG-I. TRIM25 PRYSPRY binds here, to ubiquitinate K172 in the second CARD domain with its RING domain that has E3 ligase activity. Cells with a T55I mutant of RIG-I were found to have a high permissiveness to hepatitis C virus (HCV) replication and the mutant is deficient in interaction with TRIM25, as shown by immunoprecipitation and IFN activation assays [46, 136]. The interaction of TRIM25 PRYSPRY with RIG-I CARD was tested in a surface plasmon resonance (SPR) experiment or BIAcore assay. With a curve fit of a 1:1 binding model the K_D for RIG-ICARDs wt is 1.1 μM and for

RIG-ICARDs T55I the K_D is 2.5 μM, both values lying in the same range (figure 2.41).

The SPR experiment indicates interaction between TRIM25 and RIG-I CARDs, however, the putative signaling deficient T55I, has the same affinity as the wild type in our hands. It is possible that the 1:1 kinetics model cannot be applied in this case, since the stoichiometry is not known. TRIM proteins as well as proteins containing CARD domains have been observed to assemble to highly multimeric complexes [80, 119, 127, 98, 115, 85, 13, 112]. In the SAXS experiments the CARD domains are not visible, probably due to flexibility. They might be partially folded and flexible when unbound and only get structured upon interaction. In case of folding upon binding to the partner, the kinetics would be even more complicated. Furthermore, the CARDs are labile and tend to aggregate during purification and might expose hydrophobic non-surface regions for unspecific binding to TRIM25. The interaction could also be unspecific due to unfortunate conformation of TRIM25 on the chip, but the vice-versa experiment with immobilized CARDs cannot be performed as it is not possible to immobilize the CARDs on the chip. As for the moment, we could neither establish *in vitro* ubiquitination of RIG-I with full length TRIM25 nor assemble a complex of PRYSPRY with RIG-I CARDs stable enough to be detected on native PAGE or MALLS, the work on the PRYSRPY domain is currently on hold.



A **B**

Figure 2.41: *Interaction of RIG-I CARDs and TRIM25 PRYSRPY with Surface Plasmon Resonance (SPR) (**A**) SPR interaction curves of TRIM25 PRYSPRY 434-630 with RIG-I CARDS 1-226 wt. (**B**) SPR interaction curves of TRIM25 PRYSPRY 434-630 with RIG-I CARDS 1-226 T55I.*

2.3 MDA5

The second receptor of the RLR family, MDA5, is much less studied than RIG-I. Like RIG-I, it is comprised of tandem CARDs, a helicase domain and a C-terminal domain. Atomic structures of the first part of the helicase domain of MDA5 (residues 292-490) and of the CTD are available [138, 77].

2.3.1 Protein expression and construct design

Full length MDA5 could be expressed in the baculovirus expression system in low quantities (1 mg per liter of culture). But full length protein was poorly soluble and in the final size exclusion chromatography step, a large amount of aggregates elute in the void volume of the S200/10/300 (figure 2.42). Although almost free of contaminants, the purified protein was not homogeneous on native gel and failed to crystallize. To obtain stable constructs, limited proteolysis with trypsin was performed. Based on the results of the tryptic digest, sequence alignments, secondary structure and disorder predictions, the constructs listed in figure 2.43 were designed, cloned and expressed. The two fragments with published structures are as well included in the list.

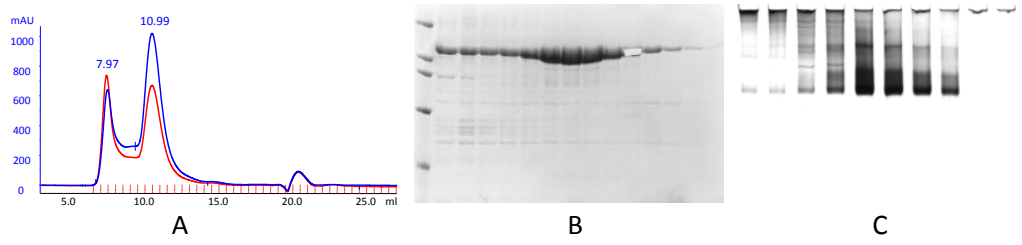


Figure 2.42: *MDA5 could be expressed and purified in low quantities. (**A**) Size exclusion profile of full length MDA5 on a S200/10/300. A large amount of soluble aggregates elutes in the void volume of the column. (**B**) The peak fractions on SDS gel. Degradation products are visible. A band was cut out to verify protein identity by mass spectroscopy. (**C**) Native gel of the same elution fractions show that the full length protein is not homogeneous.*

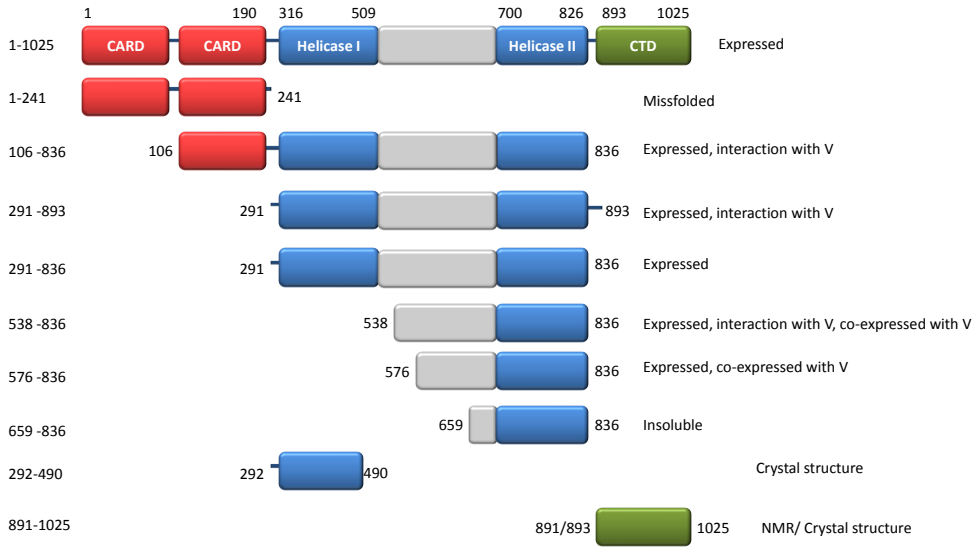


Figure 2.43: *Stable constructs of MDA5. Full length protein could be expressed in low quantities, six further constructs were identified from tryptic digest. The structure of two fragments of MDA5 was solved by other labs [138, 77].*

2.3.2 MDA5 inhibition by parainfluenza virus 5 V protein

Paramyxoviruses (*paramyxoviridae*) are a family of single-stranded negative-sense RNA viruses including paramyxoviruses (respiroviruses (e.g. sendai virus), rubulaviruses (e.g. parainfluenza virus 5 (PIV5)), morbilliviruses (e.g. measles virus), henipaviruses (e.g. hendra virus) and pneumoviruses. Paramyxoviruses use a variety of strategies to evade the IFN response or inhibit IFN signaling making use of gene products of the P/V/C gene. This single gene codes for multiple proteins due to RNA editing mechanisms and use of alternative open reading frames. The V protein inhibits IFN-β production [109]. V proteins from 13 different paramyxoviruses are able to interact with MDA5 and inhibit its function [22].

It was postulated by Childs *et al.* that a region in the C-terminus of V (corresponding to residues 104-222 in PIV5-V) inhibits activation of MDA5 by targeting a region in the helicase domain of MDA5 [22]. The minimal overlapping target sequence for all V proteins could be mapped to a fragment comprising residues 676-816 of MDA5. Interaction blocks dsRNA binding and consequent self-oligomerization of MDA5. Some V proteins

target additional regions of MDA5 (residues 287-458, 500-676 and 676-816). In contrast, none of the V proteins were able to bind to RIG-I [2, 21, 22]. Parisien *et al.* [104] tested interaction of MDA5 with V proteins of HPIV2, PIV5, measles virus, mumps virus, hendra virus and nipah virus. They identified a similar minimal binding region of 130 amino acids (residues 701-830 in MDA5, 351-479 in LGP2), which is highly homologous between LGP2 and MDA5, but not to RIG-I. Apparently, the C-terminal 60-70 amino acids of V, forming a zinc-finger, are sufficient to bind to either MDA5 or LGP2 and disrupt ATP hydrolysis activity.

The structure of parainfluenza virus V protein has been solved previously in complex with the DDB1-Cul4A ubiquitin ligase. In this context, V reprograms the ligase complex to degrade STATs and block interferon signaling [76]. Based on the factors that (i) V can be expressed in *E.coli* and has been already been crystallized and that (ii) most of our MDA5 fragments found by tryptic digest include the putative minimal binding region we attempted to form the MDA5:V complex and to crystallize it. Furthermore, since high resolution structures of the flanking parts (the first part of the helicase and the CTD) of MDA5 are available, the atomic structure of the second part of the helicase domain could serve as a basis to model the whole molecule of MDA5.

2.3.3 Expression and purification of the MDA5 : V complex

In first trials, different constructs MDA5, found by limited proteolysis, and protein V of parainfluenza virus 5 were expressed separately in insect cells and *E.coli*, respectively. The proteins were purified following the standard protocol (first nickel affinity column, TEV cleavage, second nickel affinity column) and mixed before size exclusion chromatography. Stable complexes could be formed, as for example MDA5 291-893 with protein V (figure 2.44 A, B and 2.46 A). MALLS measurements indicate a binding ratio of 1:1 (figure 2.45). However, heat shock proteins were co-expressed with all tested

constructs of MDA5 and Dnak was co-expressed with V protein. Additionally, protein V was degrading during the purification process, although degradation products did not bind to MDA5 and could be separated from the complex. After a single gel filtration the complex was not sufficiently homogeneous (figure 2.44 B). Complexes of different MDA5 constructs with full length V resisted to crystallize so far. Since the presence of chaperones and heat shock proteins often indicates problems in protein folding, we decided to co-express the two proteins in insect cells to facilitate the correct folding. An additional ion-exchange step in the protocol could improve protein purity. In parallel, limited proteolysis should determine a hopefully more stable minimal complex. MDA5 and V were mixed and digested with trypsin in a weight ratio of 1:500 protein:trypsin. The reaction was stopped after various time points and analyzed on SDS gel followed by N-terminal sequencing or mass spectroscopy. The analysis indicated several stable products of MDA5, but V protein was completely digested (figure 2.46 B). A minimal complex could not be defined from these results. Cloning and purification trials of various co-expressed constructs of MDA5 with full length V are currently being performed with a focus on shorter MDA5 fragments.

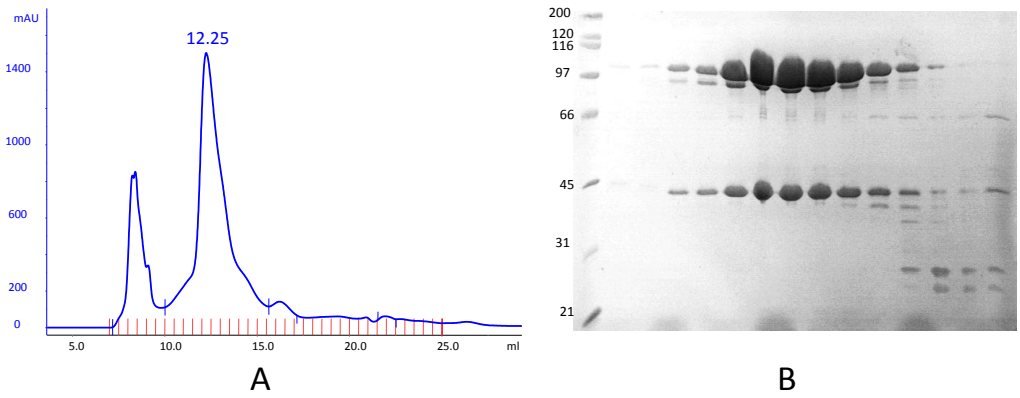


Figure 2.44: *(**A**) Size exclusion chromatography of MDA5 291-893 and V protein together. (**B**) SDS PAGE of the peak fractions of the same size exclusion chromatography. Dnak contamination (70kDa) and degradation of V protein can be observed.*

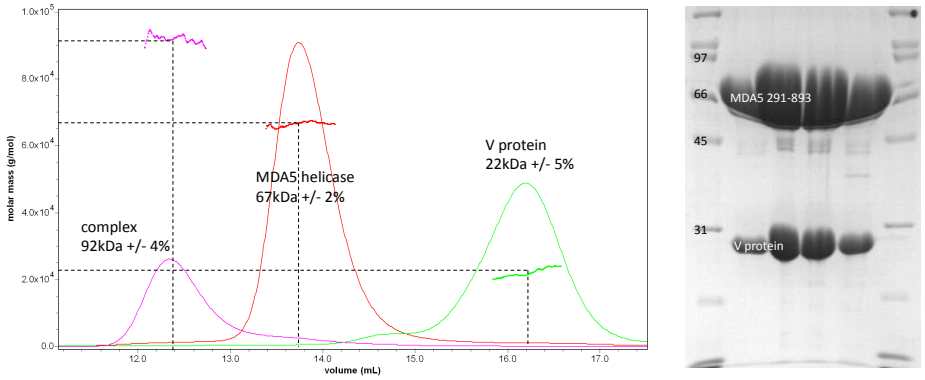


Figure 2.45: *(**A**) MALLS profile of MDA5 alone, V protein alone and the complex. (**B**) Elution fractions of the MDA5-V complex on an SDS gel*

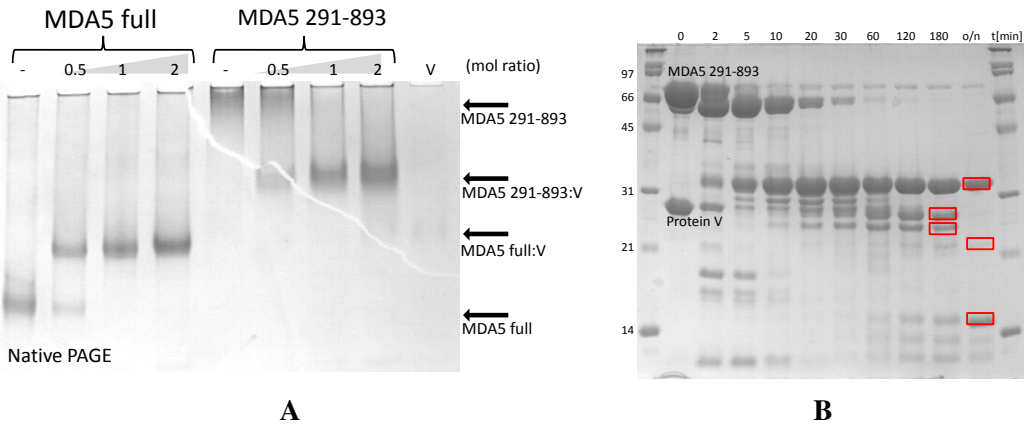


Figure 2.46: *(**A**) Native PAGE of MDA5 (full length and a construct comprising only the helicase domain 291-893) and parainfluenza virus V protein. (**B**) Results from a trypsination experiment of MDA5-V complex. Bands in red boxes were analyzed by mass spectroscopy and appeared to be fragments of MDA5.*

Chapter 3

Discussion and Future Prospects:

RIG-I

3.1 RIG-I interacts *in vitro* with different RNAs

In this thesis it has been shown that purified human RIG-I is an active RNA-dependent ATPase that binds to several T7 transcribed and purified virus derived 5'-PPP RNAs, including an influenza virus panhandle construct, and measles, rabies and ebola leaders in a 1:1 ratio. Additionally, RIG-I bound to SRP *Alu* domain RNA, a cellular RNA which does not expose its 5'-PPP in the natural context [147]. A tested tRNA did not interact with RIG-I as probably the 3'overhang at the end of the RNA stem structure, characteristic for tRNAs, might interfere with the interaction, showing in agreement with other work stems with 3'overhangs do not bind to RIG-I [123]. However, in another study, dsRNA with 5'-overhangs and 3'-overhangs interacted with RIG-I [83]. A low resolution SAXS structure of human RIG-I was modeled and the helicase domain and CTD could be localized in the envelop.

3.2 Oligomeric state of RIG-I

Our data including RNA titration, SEC-MALLS, SAXS and SANS measurements performed on several different batches of purified protein and RNA suggests a binding ratio of one molecule of RIG-I to one RNA molecule (containing one 5'-PPP). No dimerization is induced upon binding to the RNAs used in our study as it has been proposed for RIG-I [141, 154, 28, 77, 125, 120, 113].

Cui *et al.* and Schmidt *et al.* base their observations of RIG-I oligomers on gel filtration results [28, 125]. However, as shown by our SEC-MALLS results, a significant shift on an S200/10/300 does not necessarily indicate a linear augmentation of mass, as the shape of the molecule influences the migration properties on the column matrix. We also observed (data not shown) that poorly purified RNA or RNA containing palindromic sequences, self-annealed and formed heterogenic oligomers with RIG-I, containing one molecule of RIG-I bound to one to ten RNA molecules. These inhomogeneous complexes

shifted the peak even more but were not monodisperse. In the case of Cui *et al.*, the tested RVL RNA was not gel purified and by-products of the reaction might be present. Furthermore, Schmidt *et al.* and Saito *et al.* base their observation of RIG-I multimers on a band shift in native PAGE analysis [120, 125]. However a band shift does not necessarily prove dimerization, as even a small bound RNA that is highly negatively charged can significantly alter the migration properties of a protein.

A further evidence for the dimerization of RIG-I is assumed from structural work on the receptors. RIG-I and LGP2 CTDs crystallized in complex with RNA in a 2:1 protein:dsRNA ratio [77, 146, 83]. However the "dimer" might be artificially provoked by the RNA ligand that harbors two 5'-PPP ends and thus two binding sites for RIG-I. There are no protein:protein contacts between the two CTDs, while each CTD interacts extensively with the RNA. An electron microscopic (EM) structure is available that compares the monomer and dimer structures of RIG-IΔCARDs in complex with RNA [113]. The ligand RNA "dsR24" used in that study is only poorly described, probably it is a blunt-end double strand of 24 base pairs. Also in this structure, the dimer might be induced by the nature of the ligand offering two binding sites. At the moment, we are working on the production of a dsRNA carrying the 5'-PPP on each end to reproduce such a possibly"artificial" 2:1 RIG-I:dsRNA complex.

Saito *et al.* performed *in vivo* pull down studies on cells co-transfected with Flag- and Myc-tagged RIG-I and RIG-I mutants to prove self-association [120]. Figure 3.1 shows figure 3 C and D from this publication. In "C" an anti-Flag pull down of full length Myc-RIG-I was performed with Flag-tagged wt RIG-I, RIG-I CARDs, RIG-IΔCARDs and CTD. Independent of virus infection, wt Myc-RIG-I co-precipitated with all of them, interpreted as association of two independent molecules of RIG-I. But a number of points was not respected in the analysis of this experiment. Firstly, it is suspect that no clear difference is observed between the pull downs in infected and non-infected cells. Moreover,

pull-down assays with membrane associated proteins have to be carefully evaluated since many other components easily associate through the membrane to the prey. Full length RIG-I is associated to the membrane bound adapter MAVS for signaling, and IP products might represent multiple MAVS-associated molecules of RIG-I in a multi-protein complex rather than distinct soluble cytosolic RIG-I dimers. It would also be possible that multiple RIG-I molecules bind independently to one long RNA and co-precipitate.

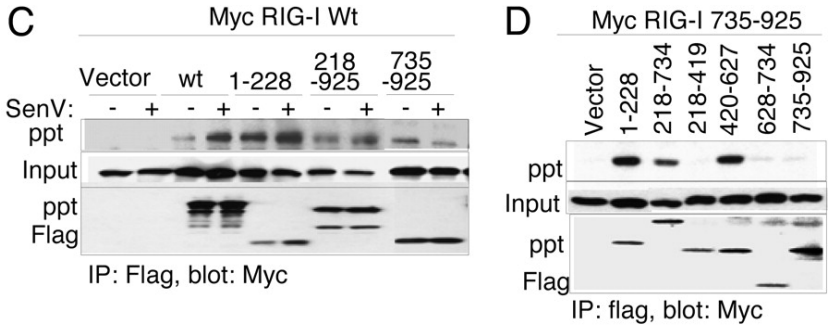


Figure 3.1: *Immunoprecipitation assays from Saito et al., Figure 3 C and D [120]. In "C" cells were co-transfected with Flag-tagged constructs of RIG-I and Myc-RIG-I wt. The Flag-tag was used for IP and the presence of wt RIG-I was detected in an anti-Myc western blot. In "D" cells were co-transfected with Flag-tagged constructs of RIG-I and Myc-CTD. IP and detection were performed as before.*

Another interesting point in this experiment is the band pattern observed for wt constructs and RIG-IΔCARDs, both showing multiple bands. The pattern for wt RIG-I could represent poly-ubiquitination but it would be astonishing that ubiquitination also took place in uninfected cells. The CARDs alone do not show such pattern.

In the same publication in figure 3 D (figure 3.1), co-immunoprecipitaition (IP) was performed to test the interaction of several RIG-I constructs with RIG-I CTD. The experiment is performed in sendai virus infected cells; results from uninfected cells are not shown. RIG-I 1-228 (CARDS), 218-734 (helicase) and 420-627 (helicase 1st part + helicase insertion) co-precipitate with the CTD, while 218-419 (helicase 1st part), 628-734 (helicase 2nd part) and the CTD itself do not. In the interpretation of the results, it was not respected that the interactions are not necessarily trans-interactions but could also be cis-

interactions reconstituting one molecule of RIG-I. Indeed, regarding the co-precipitations as intra-molecular interactions, the data would support our SAXS model (figure 2.23), in which the CTD interacts with the insertion domain of the helicase, but not with the two functional helicase domains.

Since death domains (superfamily of CARDs) are known to build up high oligomeric complexes (also known as receptor clustering) [85, 13, 80, 119, 127, 112, 156], and MAVS was shown to at oligomerize when activated [9], we assume that multimerization of RIG-I in the signaling cascade is probable. Recently, it has been proposed that the full activation of RIG-I occurs in a multi-step mechanism. First RIG-I binds RNA, then it has to be activated by ATP and then the RIG-I:RNA:ATP complex interacts with K63-linked poly-ubiquitin chains or has to be ubiquitinated to be fully active [44, 157]. The RIG-I:RNA:poly-ubiquitin complex subsequently interacts with MAVS. Theoretically, oligomerization could occur at the following steps throughout the signaling: (a) in the RIG-I:RNA complex, (b) in the RIG-I:RNA:ATP complex, (b) in RIG-I:RNA:poly-ubiquitin complex or (c) at the adapter level in the MAVS:RIG-I:RNA:poly-ubiquitin complex. Based on our data, we exclude the first two possibilities. The actual time point of the event remains to be determined. We could not yet reproduce RIG-I ubiquitination *in vitro* or interaction with K63-linked poly-ubiquitin chains *in vitro* to test weather poly-ubiquitin influences the oligomeric state of RIG-I.

3.3 Conformational changes upon RNA interaction

Saito *et al.* postulated for RIG-I a conformational change upon RNA interaction, based on a tryptic digest of full length RIG-I alone and in complex with RNA. In the complex a C-terminal fragment is stabilized [120]. We can confirm the stabilization of this C-terminal fragment upon RNA binding, but cannot interprete yet if it this is result of a local conformational change or of a simple protection of the protein by the RNA moiety from the protease activity. On basis of the SAXS and SANS experiments performed in

this study, the hypothesis of a global domain rearrangement upon RNA binding cannot be supported [120, 155]. However, the result for RIG-IΔCARDS in SEC-MALLS hints to a conformational change of this construct when bound to RNA, as it interacts with RNA but the apparent mass from MALLS is constant upon RNA binding. A SAXS experiment of this protein fragment alone and in complex with RNA should is scheduled for further investigation.

The RNA moiety cannot be clearly located in the SAXS and SANS models of the complex. This can be due to either high flexibility of the RNA or to tight attachment of the RNA along the protein molecule. The shPH RNA used in these experiments is very small (12.5 kDa) in comparison to RIG-I (106 kDa) but a longer RNA would not necessarily lead to better data, since for SAXS and SANS floppy molecules decrease the model quality. Interaction with the *Alu* domain of SRP RNA was proven by SEC-MALLS. This well structured RNA could be suitable for SAXS analysis together with RIG-I and also for further crystallization attempts. Moreover, co-crystallization with ATPase inactive mutants of RIG-I will be attempted to inhibit translocation along the RNA.

The CARD domains could not be positioned in the SAXS/SANS envelopes. They can be suspected close to the helicase moiety of the model, while the CTD is located at one tip of the Z-shaped molecule. We speculate that the CARD domains are intrinsically unfolded, that means disordered with induced fold upon interaction with a binding partner. Intrinsic folding was already observed for the CARD domain of CED-4 in complex with CED-9 [153]. In this 2:1 complex the CARD of one molecule of CED-4 is folded, while in the second molecule the CARD domain is missing because of its flexibility. Interestingly, the CARDs of RIG-I have been shown to act as inhibitors on the ATPase and translocase function [92, 28]. Results from tryptic digest of full length RIG-I support the hypothesis of poor folding, as CARDs are rapidly cleaved off the protein. They might surround the helicase domain in an unfolded state and block the access for dsRNA or

ATP. It has be shown previously, that CARD folding kinetics is complicated and that the oligomeric state of CARDs depends on pH [26, 132]. In the future, the pH dependence of RIG-I CARD folding could be further examined by circular dichroism (CD) measurements and NMR.

It was proposed that the CTD directly ”masks” the CARDs and only releases them for signaling when an RNA molecule binds [120]. Since there is no evidence from our models that CARDs and CTD interact, the activation of the CARD-mediated signaling might be more complex.

3.4 Choice of RNA constructs and purification method

Since our results differ from the common hypothesis of RNA dimerization after RNA binding, we wanted to assure that the RIG-I RNA ligands used in this study are relevant *in vivo*. For this, preliminary *in cellulo* studies were performed by Jade Louber, Gerlier lab, Lyon (figure 3.2). For the MVL construct activity was already demonstrated by Plumet and co-workers [108] . The study shows that our RNAs activate RIG-I *in vivo* and differences in IFN activation potential of the different constructs were observed. Two assays were performed, a paracrine and an autrocrine assay for IFN detection [34, 29]. In both, cells are transfected with a IFN inducing PAMP, which is the RNA in our case. Additionally, cells are infected with green fluorescent (GFP) vesicular stomatitis virus (VSV). VSV is a week virus and is inhibited in IFN treated cells, what can be measured by the GFP expression. GFP expression is thus a measure for presence of IFN secretion. Two types of cells were used: “vero” cells, sensitive to IFN but not IFN producing, and A549 cells, IFN sensible and able to secret IFNs. In the paracrine assay vero cells are incubated with the supernatant of PAMP treated A549 cells and in the autocrine assay A459 cells are co-transfected and the amount of PAMP necessary for viral inhibition is measured. The assay has still to be optimized, to distinguish weather the signal is trig-

gered by RIG-I or MDA5. This can be achieved by an siRNA approach or a cell line deficient of either receptor [34, 29].

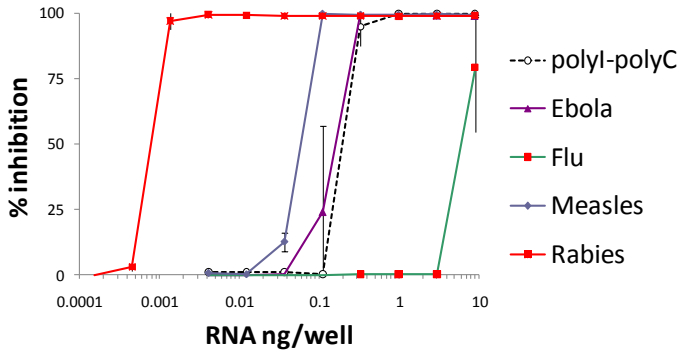


Figure 3.2: *Inhibition of VSV-GFP expression in A549 cells after transfection with T7 transcribed RNAs. Cells were oligofectamin-transfected with polyI:C, EVL, shPH, MSV, RVL. After overnight incubation they were infected with VSV-GFP at an MOI of 84. 14 hours after infection the cells GFP expression was analyzed by flow cytometrie.)*

The RNAs used in our experiments were T7 transcribed and purified by excising the band from an denaturing urea-PAGE. This purification method is sufficient for crystallographic studies, as RNAs purified with this method in our lab successfully crystallized [147]. To reach perfect homogeneity of the RNA ligand, ideally it would have to be a synthetic 5'-PPP RNA [124]. However, those are not yet commercially available. Commercially available RNAs (including siRNA) usually carry a 5' OH group. With T7 *in vitro* transcription, 5'-PPP RNA can be produced in quantities suitable for structural studies, but this method has several drawbacks. RNA polymerases (as well as DNA polymerases, p. ex. *Taq* polymerase) have the tendency to fall off the template and prematurely terminate the transcription. Additionally, they can add non-encoded nucleotides to the 3'-end of the nascent transcript. Both processes lead to rather heterogeneous 3' and 5' ends of the produced RNA and to shorter side products [88, 71, 142]. By cutting out the correct size band from a denaturing (urea) PAGE, smaller oligonucleotides can be removed. However, products differing by only several bases cannot be separated. Homogeneous ends could be achieved by the use of ribozymes [130], but since they leave cyclic phos-

phate or a 5'-hydroxyl group, respectively, at the ends of the RNA strand which might interfere with RIG-I binding, this was not tried for the moment. Although our purified RNA is sufficiently pure for binding studies on denaturing gel, native gel and MALLS, Marq *et al.* could show that for *in vivo* studies on RIG-I, the conventional gel purification and extraction is not sufficient to remove reverse-complementary by-products of the T7 polymerase, transcribed by self-priming from the newly synthesized RNA as template [96, 124, 17]. Both, product and by-product can re-anneal after purification and induce a RIG-I dependent interferon response *in vivo*, even if they represent only a small fraction (1-2%) of the purified RNA. The *in vivo* results on our RNAs and the differences between the constructs be might artifacts based on these contaminating double strands. Different levels of contamination in the different batches could lead to the differences in interferon response. All publications including *in vivo* data on RIG-I with T7 transcribed RNA have to be carefully revised in the light of possible dsRNA contamination. To prove *in vivo* relevance of our RNAs, a purification protocol has to be established, removing all by-products of T7 polymerase. Alternatively synthetic 5'-PPP-RNA could be used.

3.4.1 Mallard duck RIG-I structure

For which purpose is RIG-I an ATPase? Why does it act as a translocase? How does static 5'-PPP binding go together with translocating activity? And does the helicase contribute to dsRNA recognition? To answer these questions it is important to achieve high resolution structures of the full length receptor alone and in complex with RNA. The mallard duck RIG-I structure will provide important details for modeling the human structure into the existing SAXS and SANS envelopes. Hopefully it will be possible to localize the CARDs for the first time. The structure will also the basis for mutagenesis studies and protein engineering to design a crystallization prone version of the human protein.

Chapter 4

Article Manuscript

This chapter is confidential.

Structural analysis of metal chelating inhibitor and mononucleotide binding to influenza H1N1 (2009) polymerase PA endonuclease domain

Eva Kowalinski[1,2] and Stephen Cusack[1,2*]

[1]European Molecular Biology Laboratory, Grenoble Outstation 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France.

[2]Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France.

*Corresponding author:
Stephen Cusack
Tel. (33)476207238
Email: cusack@embl.fr

To be submitted.

Abstract

The ongoing potential threat to humans of highly pathogenic H5N1 influenza viruses and the 2009 H1N1 influenza pandemic has highlighted the need for alternative therapeutic options to treat influenza infections which cannot be prevented by vaccination. Recent structural studies on functional domains of the influenza polymerase open the way to a structure based approach to optimise inhibitors of viral replication that directly target the polymerase. In particular the unique cap-snatching mechanism of viral transcription can be inhibited by targeting either the PB2 cap-binding domain or PA endonuclease domain. Here we describe crystal structures of the pandemic H1N1 PA endonuclease domain with a series of diketo acid inhibitors all of which chelate the two manganese ions bound in the active site of the protein. Comparison of the mode of binding of the different compounds and with that of a mononucleotide phosphate highlights firstly, the plasticity of the active site and and secondly, shows how different substituent groups on the same basic metal binding scaffold can be orientated to bind in distinct sub-pockets within the active site cavity. These structures will be useful in optimising the design of inhibitors targeting the cap-snatching activity of influenza virus polymerase.

1. Introduction.

Influenza virus replicates in the nucleus of infected cells where the heterotrimeric viral RNA-dependent RNA polymerase is responsible for replication and transcription of the segmented single-stranded, negative sense RNA genome (vRNA). Transcription of viral mRNAs occurs through a unusual 'cap-snatching' mechanism (Plotch et al, 1981) which has only been reported for the genus of negative strand segmented viruses, including orthomyxoviruses (notably influenza), bunyaviruses and arenaviruses. Cap-snatching involves the binding of host cell pre-mRNAs via their 5′ cap structure to the PB2 subunit of the polymerase followed by cleavage at nucleotides 10-13 by an endonuclease activity which resides in the PA subunit of the polymerase. The short capped oligomers then serve as primers for transcription of the viral mRNAs by the PB1 subunit of the polymerase. The viral transcripts are poly-adenylated by a stuttering mechanism at a conserved U-rich region of the template vRNA (Poon et al, 1999); thus the viral mRNAs have both the 5′ and 3′ signals to be competent for translation after nucleo-cytoplasmic export. In the last two years, crystal structures of the two functional domains involved in cap-snatching have been determined (reviewed in (Ruigrok et al, 2010)). The cap-binding domain resides in the central region of the PB2 subunit and has a unique fold while still binding the m^7G ligand by means of an aromatic sandwich, similar to other cap-binding proteins (Guilligay et al, 2008). The endonuclease domain is at the N-terminus of the PA subunit and has a core fold similar to other two-metal dependent nucleases of the PD... DxK family (Dias et al, 2009; Yuan et al, 2009). Indeed, the isolated, recombinant endonuclease domain has divalent cation dependent, *in vitro* nuclease activity with a strong preference for manganese ions, consistent with the much tighter binding of manganese than magnesium (Crepin et al, 2010). Since transcription by cap-snatching is essential for virus replication, inhibition of either the cap-binding, endonuclease or polymerase activities, are potential means of anti-viral therapy and indeed each of these targets have been and are being actively pursued (Das et al; Gong et al, 2009; Krug & Aramini, 2009). Here we exploit the availability of endonuclease domain crystals to provide the first detailed structural information on inhibitor binding to the influenza polymerase, which will be useful for further optimisation of these drug leads.

The need for new therapeutic options targeting influenza virus is now widely recognised following recent serious events, such as the circulation of highly pathogenic avian H5N1 strains and the 2009 H1N1 pandemic, which highlighted the vulnerability of the world

population to novel influenza viruses for which there are no vaccines and limited, resistance prone anti-viral drugs.

Merck identified a number of 4-substituted 2,4-dioxobutanoic acid compounds that specifically inhibit influenza polymerase endonuclease activity with IC50s in the range 0.2 to 29 μM (Hastings et al, 1996; Nakazawa et al, 2008; Tomassini et al, 1994). They also reported that a substituted 2,6-diketopiperazine natural compound (Flutimide) from the fungus *Delitschia confertaspora* and its derivatives showed endonuclease inhibition activity and inhibition of influenza A and B virus replication in cell culture (Tomassini et al, 1996). Bristol-Myers Squibb identified N-hydroxamic acid and N-hydroxyimide compounds that inhibit the endonuclease (Cianci C, 1996). Roche discovered a new class of endonuclease inhibitors, with IC50s down to 3 μM based on considerations of the likely divalent cation binding properties of the enzyme (Parkes et al, 2003). More recently other compounds have been shown to inhibit the endonuclease, for instance green tea catechins (Kuzuhara et al, 2009) or phenethylphenylphthalimide analogs derived from thalidomide (Iwai et al). Whilst none of these endonuclease inhibitors appear to have entered preclinical development, it is interesting to note that the HIV integrase inhibitor Raltegravir (also a diketobutanoic acid derivative) targets a two cation containing active site with some similarities to the influenza endonuclease (Hare et al; Summa et al, 2008).

The X-ray structure of the A/Victoria/3/1975(H3N2) endonuclease (Dias et al, 2009) was obtained with a crystal form in which crystal contacts blocked the access to the active site. In particular, Glu59, from a neighbouring molecule in the asymmetric unit binds to one of the metals in the active site (Supplementary Fig 4 in (Dias et al, 2009)). Despite many trials, no structures have been obtained of H3N2 endonuclease with bound inhibitors or substrate or product analogues, severely limiting the possibility for structure-based inhibitor optimisation. To overcome this block, we investigated the possibility that the endonuclease from other influenza strains having slightly different amino acid sequences, might yield more useful crystal forms. We found that a construct comprising residues 1-198 of A/California/04/2009(H1N1) (2009 pandemic strain), expressed from a synthetic gene, readily crystallised with and without relevant ligands. In the residue range 1-198 of the construct, the H1N1 sequence differs in 11 positions from that of H3N2; however the active site and close vicinity does not differ between the two strains. In particular structures have been determined of the A/California/04/2009-H1N1 influenza endonuclease complexed with two divalent metal ions and four 4-substituted 2,4-dioxobutanoic acid inhibitors (Tomassini et al, 1994). The compounds are 2,4-dioxo-4-phenylbutanoic acid (DPBA), 4-[N-benzyl-3-(4-

chlorobenzyl)piperidin-3-yl]2,4-dioxobutanoic acid (denoted EMBL-R05-03) and 4-[1-cyclohexylmethyl-4-(p-chlorobenzyl)piperidin-4-yl]2,4-dioxobutanoic acid (denoted EMBL-05-02) and 4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid (denoted EMBL-R05-01) (Table 1). These compounds are reported to have IC50s of 3.7, 1.1, 0.33 and 0.19 μM (Tomassini et al, 1994). Two additional structures are presented of the A/California/04/2009(H1N1) influenza endonuclease with bound mononucleotides rUMP and dTMP, components of the RNA (and DNA) endonuclease substrate. In a previously reported structure of rUMP bound to the influenza nuclease, the critical manganese ion in site 1 is missing (there is only a magnesium in the site 2) and the nucleotide is poorly ordered (Zhao et al, 2009). In our mononucleotide structures both metal ions are present and the ligands are perfectly ordered.

All these structures show in detail how these compounds bind directly to the metal ions as well as interacting with a number of residues in the active site. Furthermore several of the interacting residues change conformation upon ligand binding, information which was unavailable before. This three-dimensional knowledge of the ligand interacting residues and the regions of plasticity of the active site is critical for the optimised design of modifications to existing inhibitors to improve their potency or for structure based design and optimisation of novel inhibitors that effectively block endonuclease activity.

2. Results.

2.1 Comparison of PA-Nter from H1N1, H5N1 and H3N2 structures.

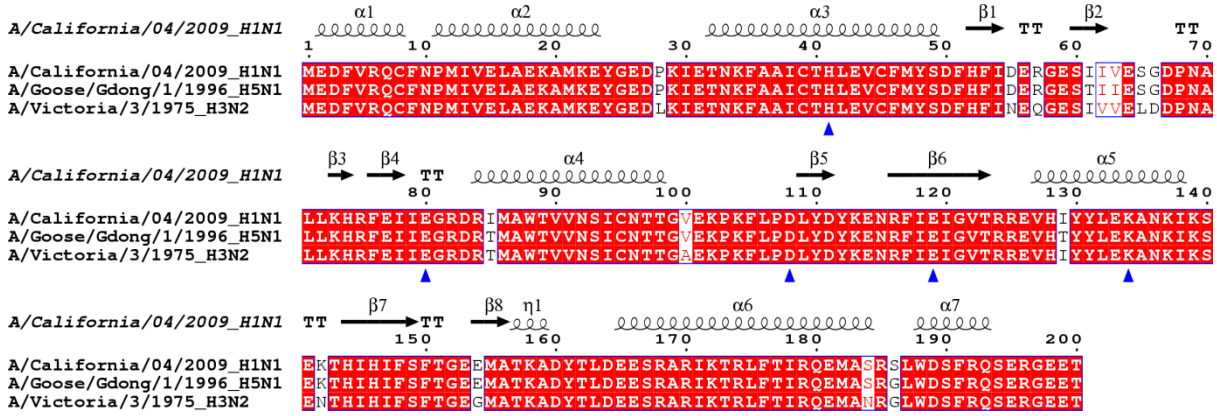


Figure 1. Sequence alignment of H1N1, H5N1 and H3N2 PA-Nter sequences for which crystal structures are known. Blue triangles indicate conserved cation binding (His41, Glu80, Asp108, Glu119) and catalytic (Lys134) residues**. The secondary structure of the H1N1 domain is shown over the alignment.**

There are respectively 5 and 12 differences between the H1N1 sequence compared to the H5N1 and H3N2 sequences (Figure 1). The serine at position 186 of A/California/04/2009(H1N1) is unique. It is not known whether these differences play a role in inter-species transmission or virulence. The most variable region is between residues 55-66 which forms part of a mobile inserted element (residues 53-73), absent in influenza C strains, that is solvent exposed and usually disordered. In the H3N2 and H1N1 structures this element is well defined in some chains in the asymmetric unit and shows a significantly different orientation between the two strains (Figure 2). However this is likely due to different crystal contacts since in the H3N2 structure, Glu59 within the loop, interacts with a divalent cation in the active site of a neighbouring molecule. The function of this element is unknown. Apart from this, the structures are overall very similar, with small differences in helix orientation and loop conformation (Figure 2).

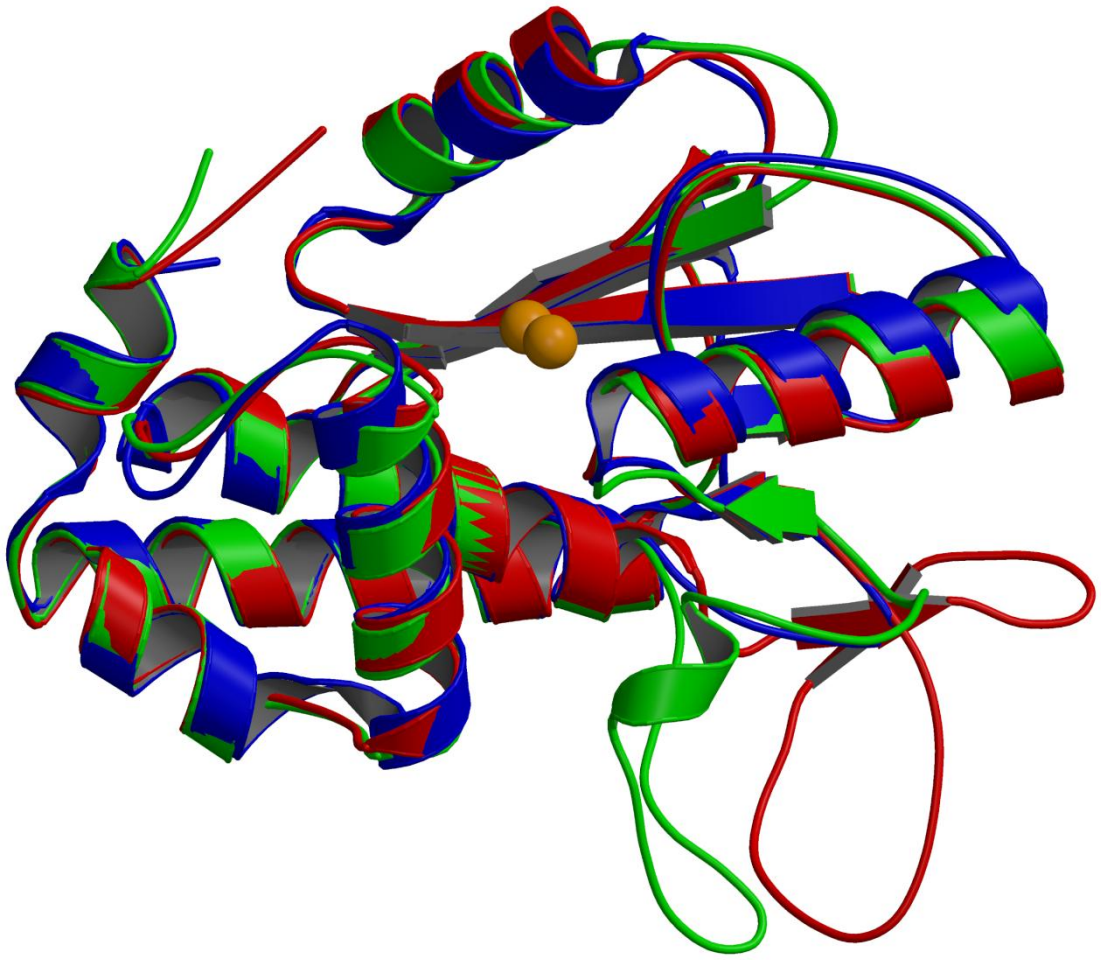


Figure 2. Superposition of PA endonuclease structure from H3N2 (green, PDB entry 2W69), H5N1 (blue, PDB entry 3EBJ) and H1N1 (red, this work). Bound Mn^{2+} ions are represented by orange spheres. Flexible region 53-73 is at the bottom right and only ordered in certain chains from the H3N2 (B chain) and H1N1 (D chain of dTMP complex) structures. For H5N1, region 53-73 is not visible.

2.2 Divalent cation binding in native structure

Previous studies have shown that there are two divalent cation binding sites in PA-Nter: site 1 is co-ordinated by His41, Asp108, Glu119 and the carbonyl oxygen of Ile120 and site 2 by Glu80 and Asp108 (Dias et al, 2009) (Figure 3). Further *in vitro* studies show that manganese ions bind strongly and exclusively to site 1 but that both manganese and magnesium ions can bind at site 2 (Crepin et al, 2010; Dias et al, 2009; Yuan et al, 2009; Zhao et al, 2009). These observations on the isolated domain are consistent with earlier results on the metal dependent endonuclease activity within native RNPs (Doan et al, 1999). As the exact *in vivo* situation is unknown, all subsequent work, notably crystallization, was performed in the presence of 2 mM $MnCl_2$ and 2 mM $MgCl_2$. Under these conditions, two

divalent cations are observed in the active site in the native, unliganded H1N1 structure. Site 1 is clearly identified as a manganese ion by strong anomalous scattering observed in a dataset measured at a wavelength of 1.55 Å (Mn peaks between 7.4 and 10.1σ) (at this wavelength sulphur anomalous peaks due to the sulphurs of most methionine and cysteine sulphurs are clearly visible above 3σ in the anomalous scattering maps). On the other hand site 2 is probably occupied by a magnesium ion, since there is no anomalous peak and refinement with a fully occupied magnesium gives B-factors similar to that of a fully occupied manganese in site 1. However, a partially occupied or more thermally mobile manganese cannot be ruled out, In any case, both sites have octahedral co-ordination: site 1 by His41, Asp108, Glu119, Ile120 and two water molecules (W4 and W5) and site 2 by Glu80 and Asp108 and four water molecules (W1, W2, W3 and W4) (Figure 3).

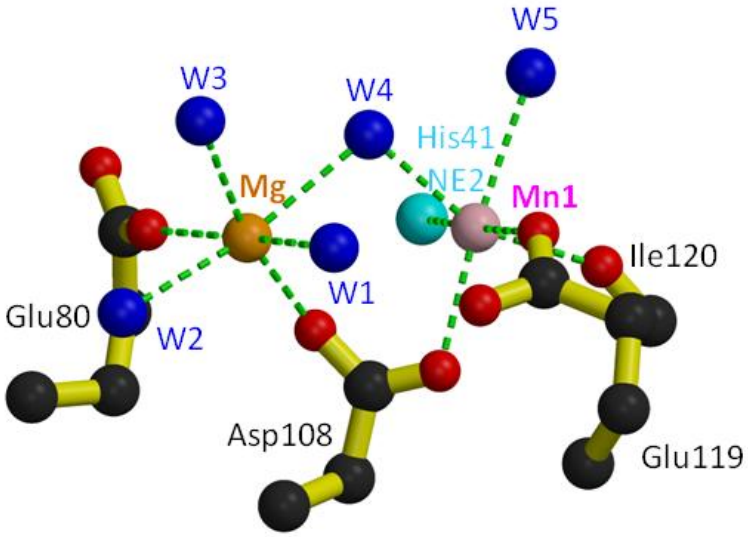


Figure 3: Divalent ion co-ordination in the native endonuclease structure. Manganese ions are pink spheres and co-ordinating water molecule blue spheres and the ion co-ordination is shown with green lines. His41 NE2 is represented by a cyan sphere.

2.3 Thermal stabilization of PA-Nter due to diketoacid inhibitor binding.

The thermal stability of the endonuclease in the presence of the diketoacid inhibitors was tested by Thermofluor assays in which a hydrophobic fluorophore has little affinity for native proteins but binds to denatured proteins, leading to an increase of fluorescence (Ericsson et al, 2006). The apparent melting temperature (Tm) of denaturation can be

obtained from the temperature dependence of the fluorescence (Figure 4). The estimated Tm is 53.5, 65, 69, 71 and 69 °C for respectively no ligand, DPBA, EMBL-R05-01, EMBL-R05-02 and EMBL-R05-03. This confirms that EMBL-R05-01, EMBL-R05-02 and EMBL-R05-03 all chelate the cations in the active site of the endonuclease but enhance the thermal stability even more than for DPDA (Crepin et al, 2010; Dias et al, 2009).

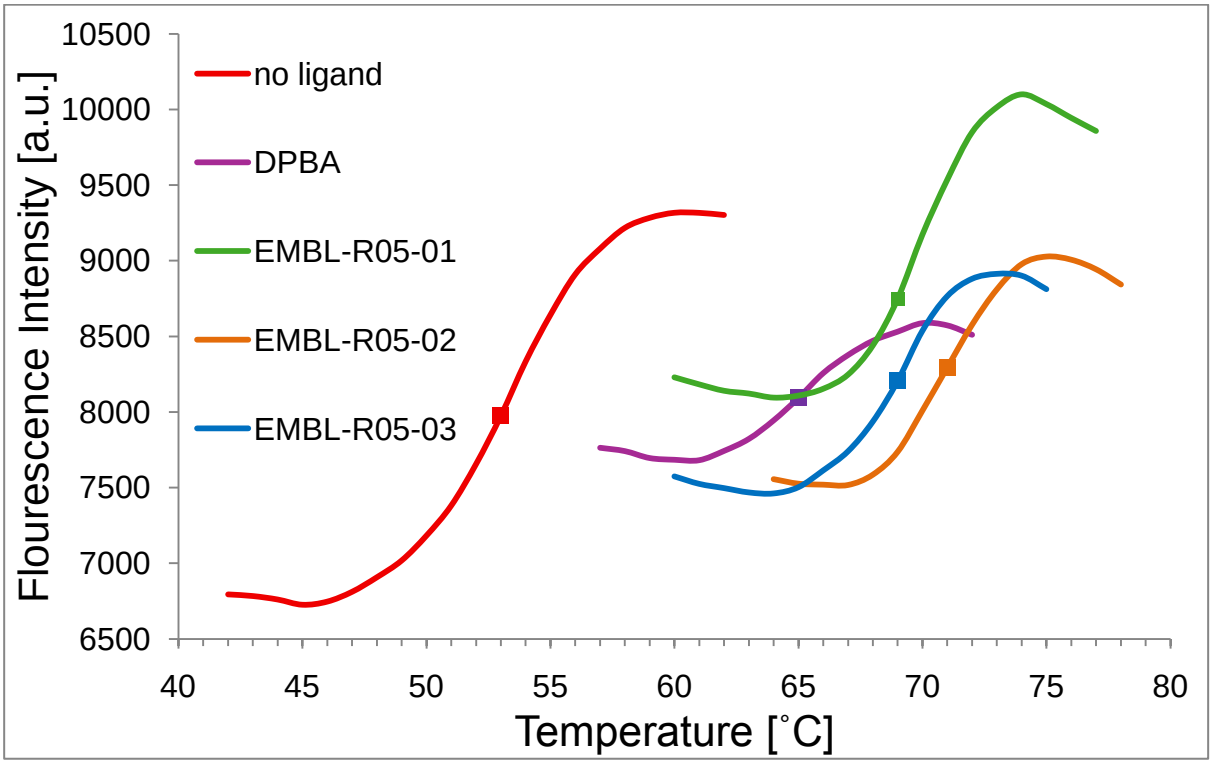


Figure 4. Thermal shift assays were performed with 5 µM H1N1 PA-Nter in 100 mM Hepes pH 7.5, 100 mM NaCl, 1 mM $MnCl_2$, 1 mM $MgCl_2$, 1 mM DTT in the presence of absence of 500 µM of the indicated inhibitors and a 5× dilution of SYPRO Orange dye (Invitrogen) as described. The dye was excited at 490 nm and the emission light was recorded at 575 nm while the temperature was increased by increments of 1°C per minute from 45-93 °C (25 to 73 °C for no ligand). The estimated Tm's are 53.5, 65, 69, 71 and 69 °C for respectively no ligand, DPBA, EMBL-R05-01, EMBL-R05-02 and EMBL-R05-03.

2.4 DPBA bound structure

The structure was determined at 2.3 Å resolution. Electron density for the ligand is shown in Supplementary Figure 1; the phenyl ring is less well-defined due to rotational flexibility as it is not conformational stabilized by any other interactions. The two metal co-

ordination is shown in Figure 5. Compared to the unbound state (Figure 3), metal ion co-ordinating water molecules W3, W4 and W5 are replaced by oxygens from the ligand. An identical configuration was observed for DPBA binding in the active site of bunyavirus cap-snatching endonuclease (Reguera et al, 2010).

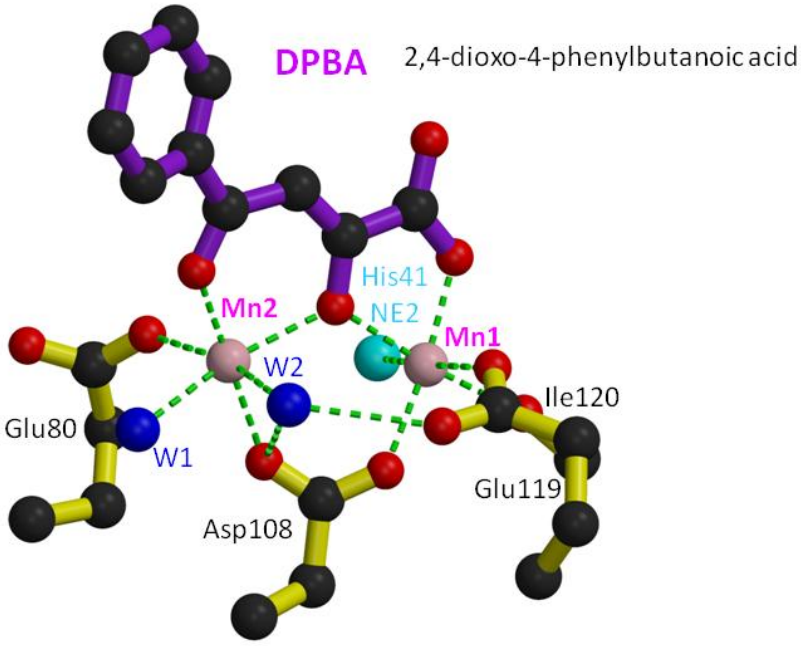


Figure 5: Top. Divalent ion co-ordination in the DPBA bound structure. Manganese ions are pink spheres and co-ordinating water molecule blue spheres and the ion co-ordination is shown with green lines. His41 NE2 is represented by a cyan sphere.

2.5 EMBL-R05-03 bound structure

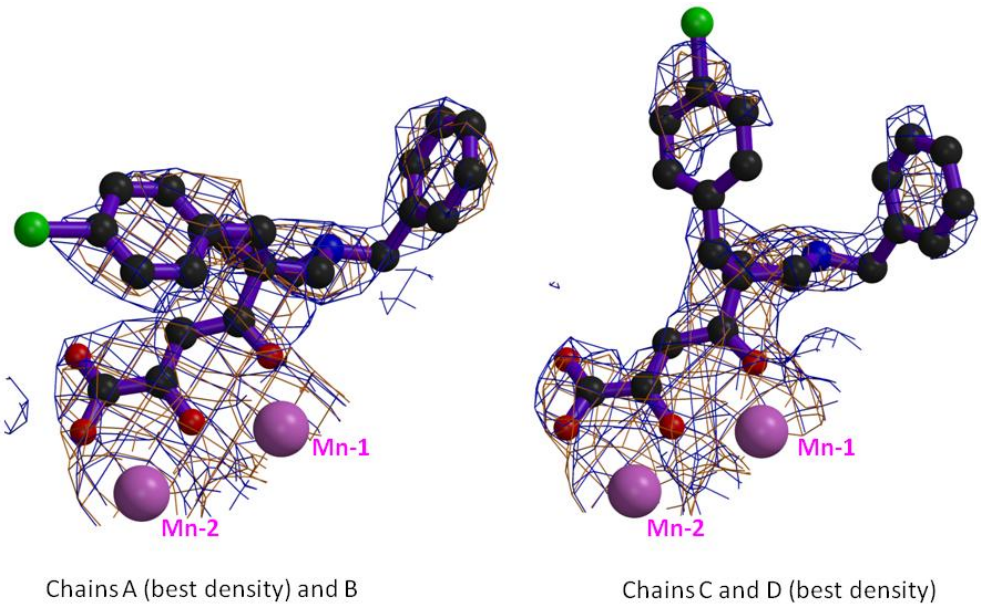


Figure 6: Two different conformations of bound EMBL-R05-03 in respectively the A, B and C, D chains in the asymmetric unit. Blue contour: final 2Fo-Fc electron density at 1.0σ, orange contour Fo-Fc omit map at 2.8σ. Manganese ions are pink spheres.

As shown in Figures 6 and 7, EMBL-R05-03 is observed in two different conformations in respectively chains A, B (denoted conformation 3A) and C, D (denoted conformation 3D) in the asymmetric unit. The electron density for the metal binding moiety of the ligand is significantly stronger than for the 'arms' presumably due to flexibility of the latter. Conformation 3A has significantly better electron density than 3D.

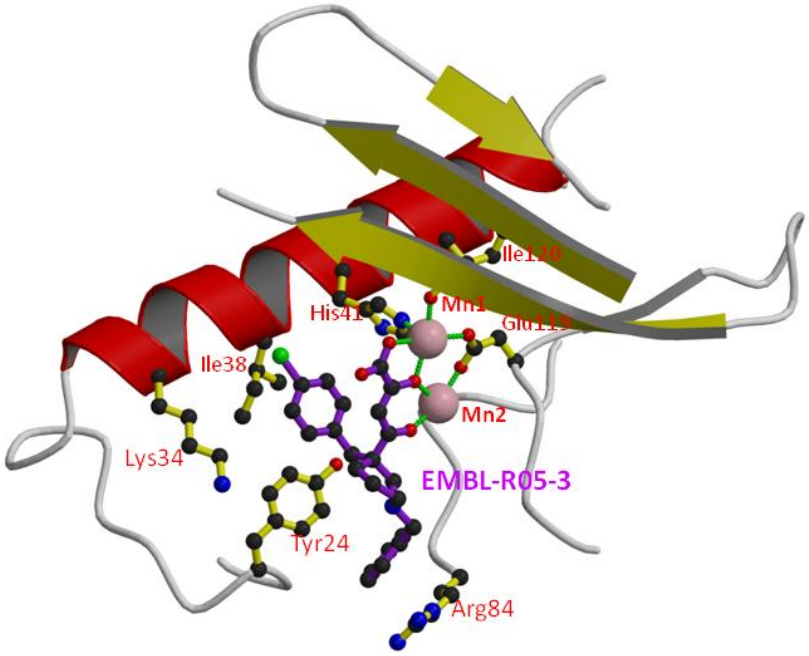


Figure 7a: H1N1 endonuclease active site showing the bound EMBL-R05-3 in conformation 3A, the two manganese ions and key active site residues that interact with the compound (notably Arg84) or are close to it. Manganese ions are pink spheres and the ion co-ordination is shown with green lines.

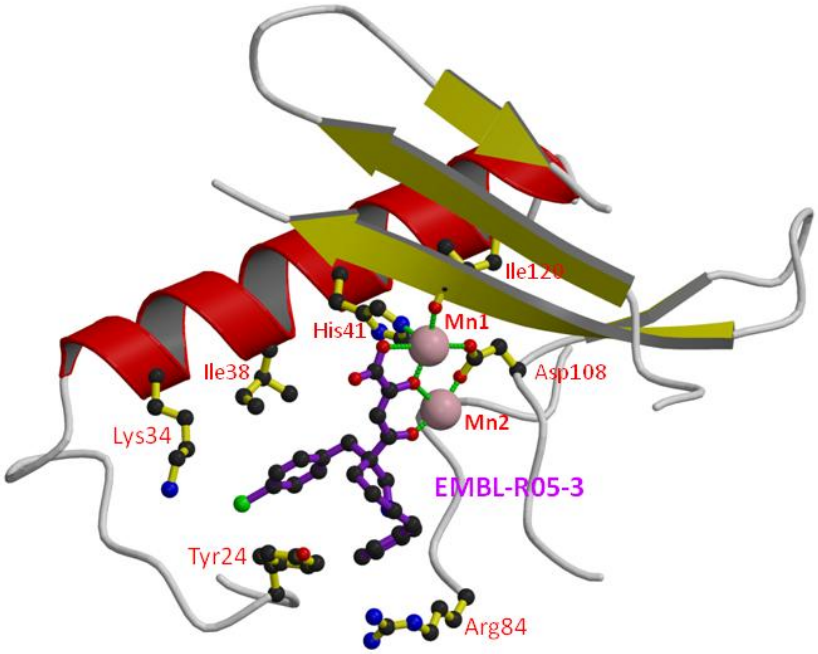


Figure 7b: H1N1 endonuclease active site showing the bound EMBL-R05-3 in conformation 3D, the two manganese ions and key active site residues that interact with the compound (notably Tyr24 and Arg84)) or are close to it. Manganese ions are pink spheres and the ion co-ordination is shown with green lines.

2.6 EMBL-R05-02 bound structure

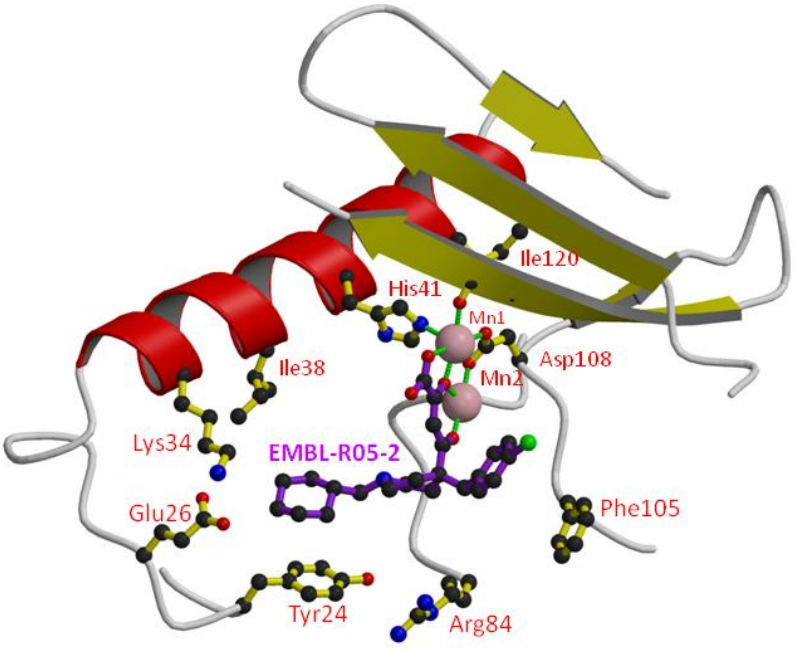


Figure 8. Bound EMBL-R05-2, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Phe105) or are close to it. Manganese ions are pink spheres and the ion co-ordination is shown with green lines.

2.7 EMBL-R05-01 bound structure

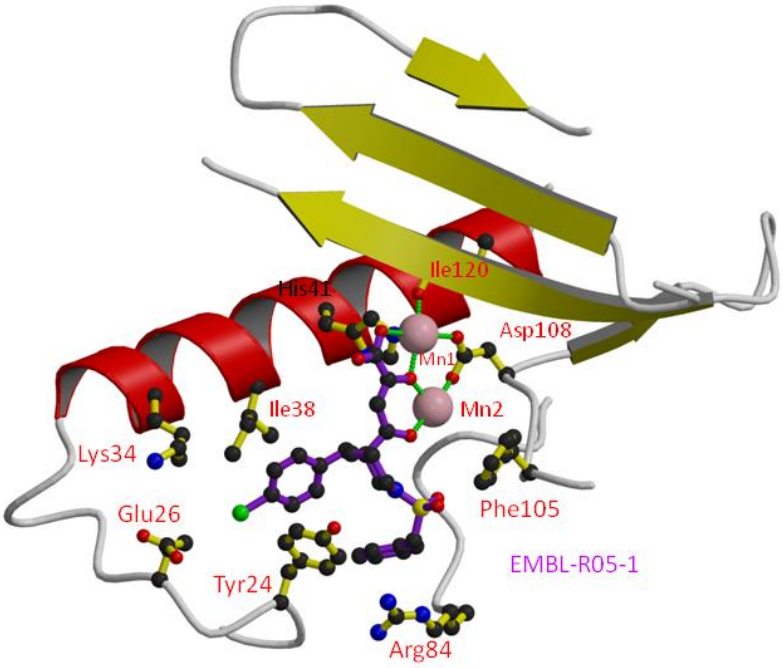


Figure 9. Bound EMBL-R05-1, the divalent cations (two manganese) and key active site residues that interact with the compound (notably Tyr24 and Arg84) or are close to it. Manganese ions are pink spheres and the ion co-ordination is shown with green lines.

2.8 rUMP bound structure

Co-crystallisation trials were attempted with all four mononucleotides, deoxy and oxy. The only compounds that resulted in a structure were dTMP and rUMP, both of which gave large, well-ordered crystals in a new orthorhombic space-group (Supplementary Figure 5). Apart from the obvious differences in the ribose and base, the two structures are essentially the same (Supplementary Figs 4 and 5) so only the rUMP structure will be described. As shown in Figure 10, the rUMP binds with two oxygens of the phosphate completing the co-ordination sphere of Mn1, one of them also co-ordinating Mn2. The base is well stacked on Tyr24 and Lys34 makes a hydrogen bond to the O2 position. The ribose

hydroxyl groups do not make hydrogen bonds to the protein, consistent with the fact that deoxy ribose binds equally well and the protein is a DNAase as much as an RNAase (Dias et al, 2009). The conformation we observe for rUMP is quite different from that previously published (PDB entry 3hw3 (Zhao et al, 2009)). The latter structure was obtained by soaking nucleotides into existing crystals of the endonuclease in the absence of manganese and the electron density is very poor. In this structure, a water molecule replaces Mn1 and a magnesium ion replaces Mn2 (Figure 11). This difference in metal ligation is reflected in the altered conformation of Glu119. The ribose and base positions are quite different from in our structure (Figure 10) and unable to interact with Lys34 or Tyr24. We suspect that the differences between the two structures reflect firstly the lack of manganese and secondly the fact that soaking pre-grown crystals does not allow the active site to adapt to the ligand as is more likely the case for co-crystallisation.

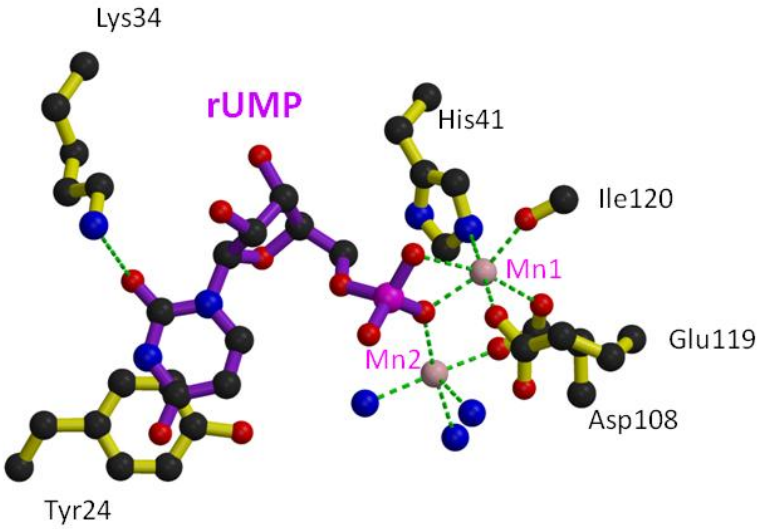


Figure 10. Bound rUMP showing stacking of the base on Tyr24 and hydrogen bonding to Lys34. Manganese ions are pink spheres, water molecules as blue spheres and the ion co-ordination is shown with green dotted lines.

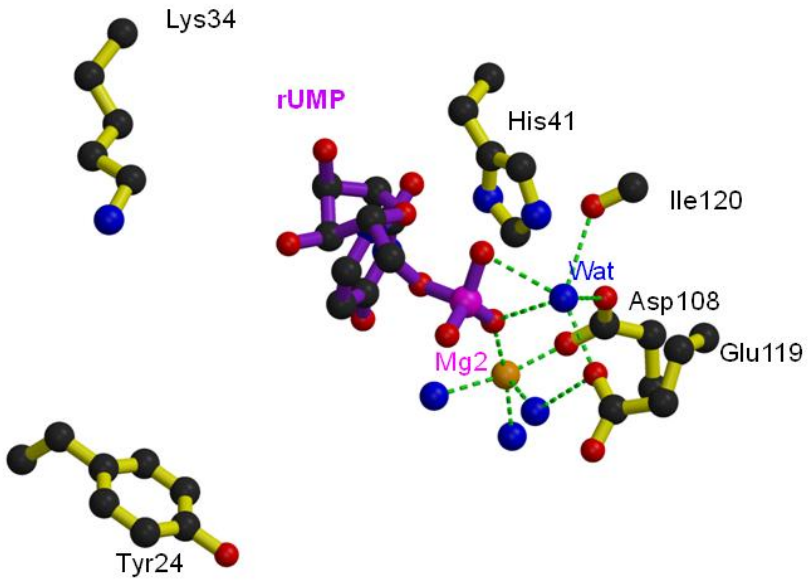


Figure 11. **H5N1 PA with bound rUMP as drawn from PDB entry 3hw3** (Zhao et al, 2009) with the protein in the same orientation as Figure 10. In this structure, a water molecule replaces Mn1 and a magnesium ion replaces Mn2.

3. Discussion

3.1 Comparison of binding mode of three diketo acid inhibitors.

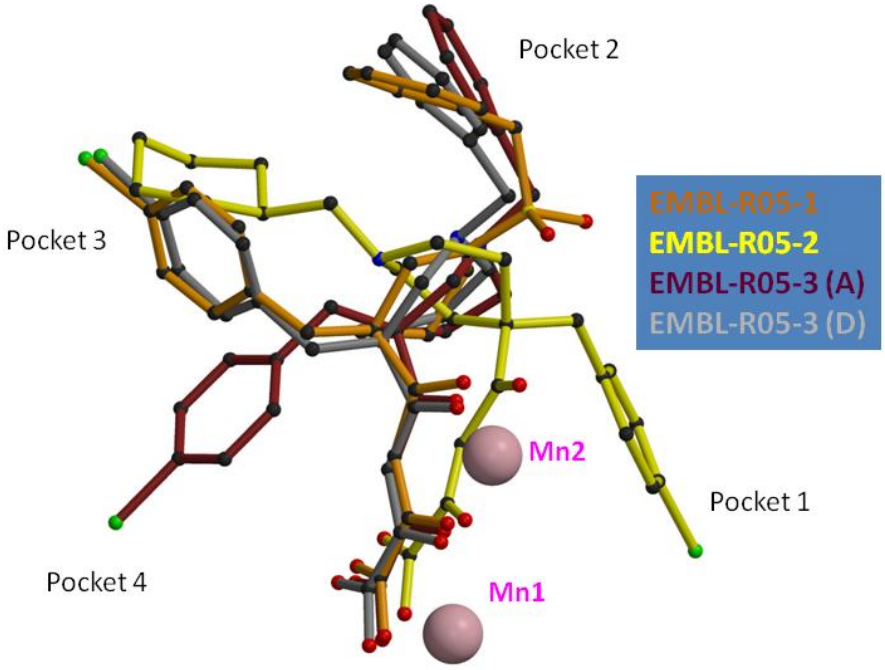


Figure 12. Superposition of all diketo inhibitor compounds bound in PA active site.

As shown in Figure 12, the mode of binding of the three diketo inhibitors to the metals is conserved (although there is some variability in exact position) but the two ‘arms’ of each compound are inserted into different combinations of the pockets 1 to 4. EMBL-R05-1 has a similar configuration to EMBL-R05-3D, with the two arms occupying pockets 2 and 3. EMBL-R05-3A occupies pockets 2 and 4. EMBL-R05-2, which differs notably from R05-1 and R05-3 in the point of substitution on the piperidinyl ring (Table 1) occupies pocket 3 and uniquely pocket 1 (Figure 13). More potent and specific compounds could be perhaps designed that occupy more than two of the pockets.

3.2 Plasticity and induced fit of the active site upon ligand binding.

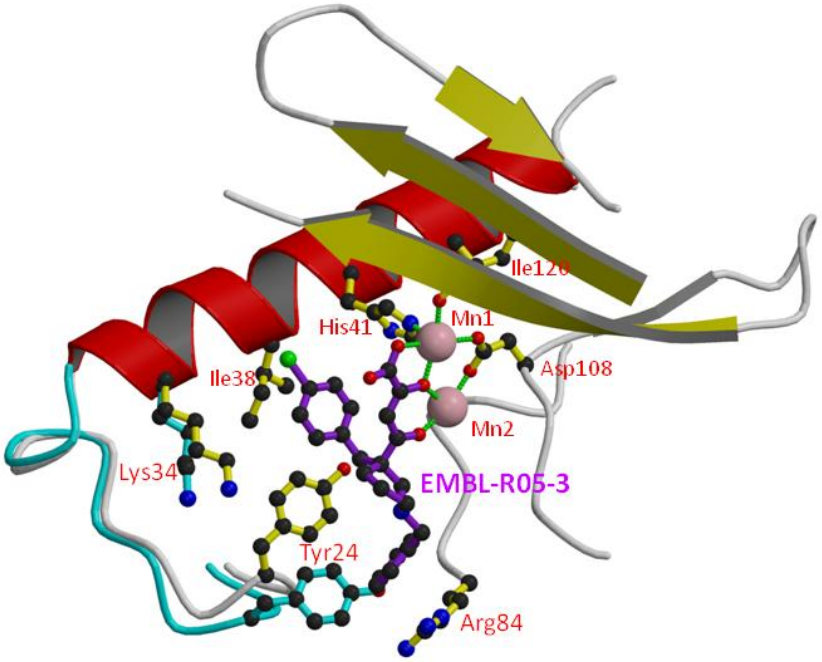


Figure 13: **Diagram comparing the native (cyan) and EMBL-R05-3 bound form in the better defined conformation 3A**. Tyr24 side-chain moves to partially stack with the chlorobenzene of EMBL-R05-3. Manganese ions are pink spheres and the ion co-ordination is shown with green lines.

Figure 13 illustrates the conformational changes and ordering that occur upon EMBL-R05-03 binding, in particularly of the loop around Tyr24, which is poorly ordered in the native structure. Indeed Tyr24 is observed to be in three rotamers depending on which of the four

pockets are occupied (Figure 14). Side-chains of other residues (e.g. Arg84, Lys34, Glu26, Phe105) also change depending on which pockets the ligand is occupying. This indicates a plasticity of the active site and an induced fit mode of ligand binding. An important conclusion for designing more potent inhibitors is to ensure that the extensions ('arms') to any ion-binding scaffold optimise interactions in one or more pocket. Imperfect matching will lead to residual flexibility and sub-optimal potency, as seems to be the case for the current compounds, none of which exhibit very well defined, full occupancy binding modes.

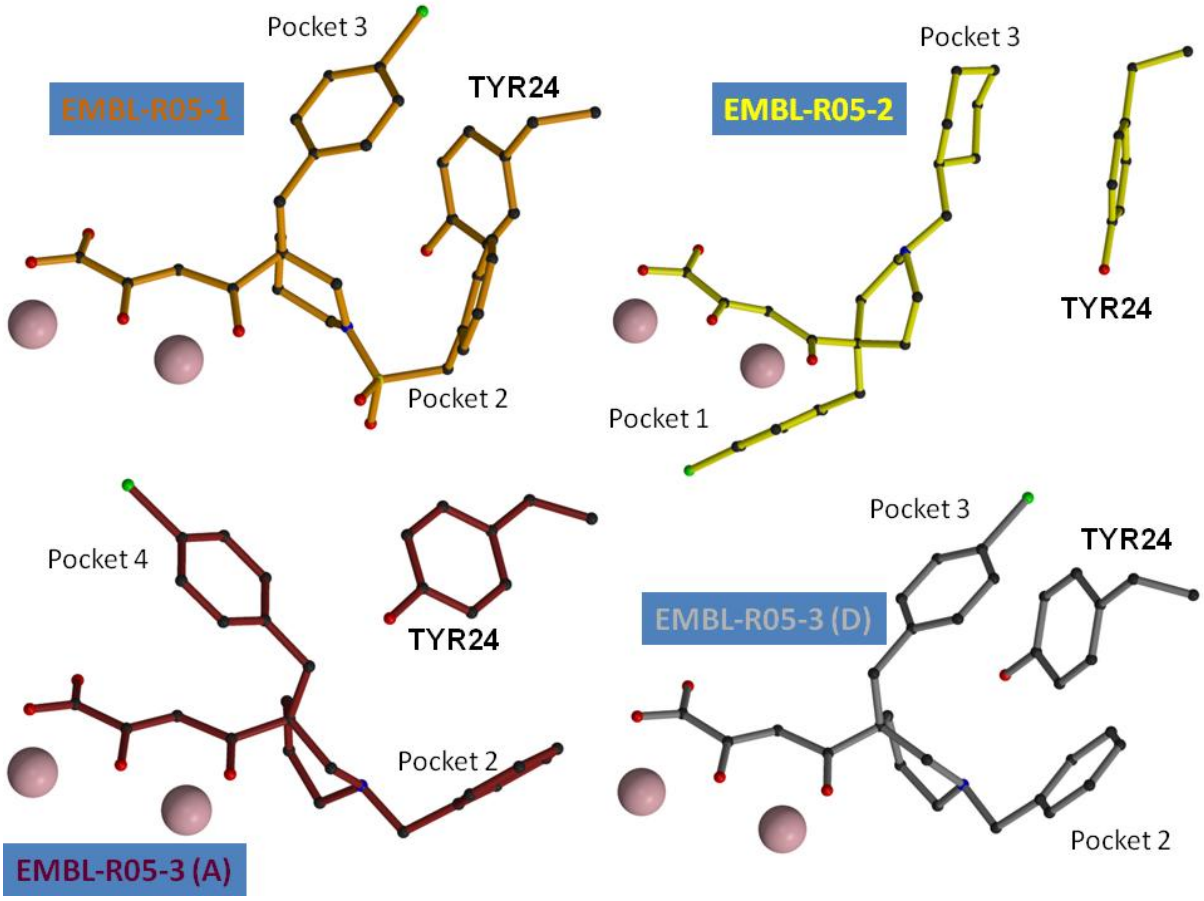


Figure 14. Configuration of the diketo inhibitor compounds with respect to residue Tyr24.

3.2 Biological significance of the rUMP binding site.

Unlike the diketo inhibitors, the binding of dTMP/rUMP exhibits a very well defined, full occupancy binding mode. From Figure 15, which compares the binding of dTMP and EMBL-R05-3A, it can be seen that the ribose of the mononucleotide is in pocket 4 and the base,

perfectly stacking on Tyr24, is in pocket 3. The apparent optimisation of this binding might reflect its biological significance as representing a part of the natural nucleic acid substrate.

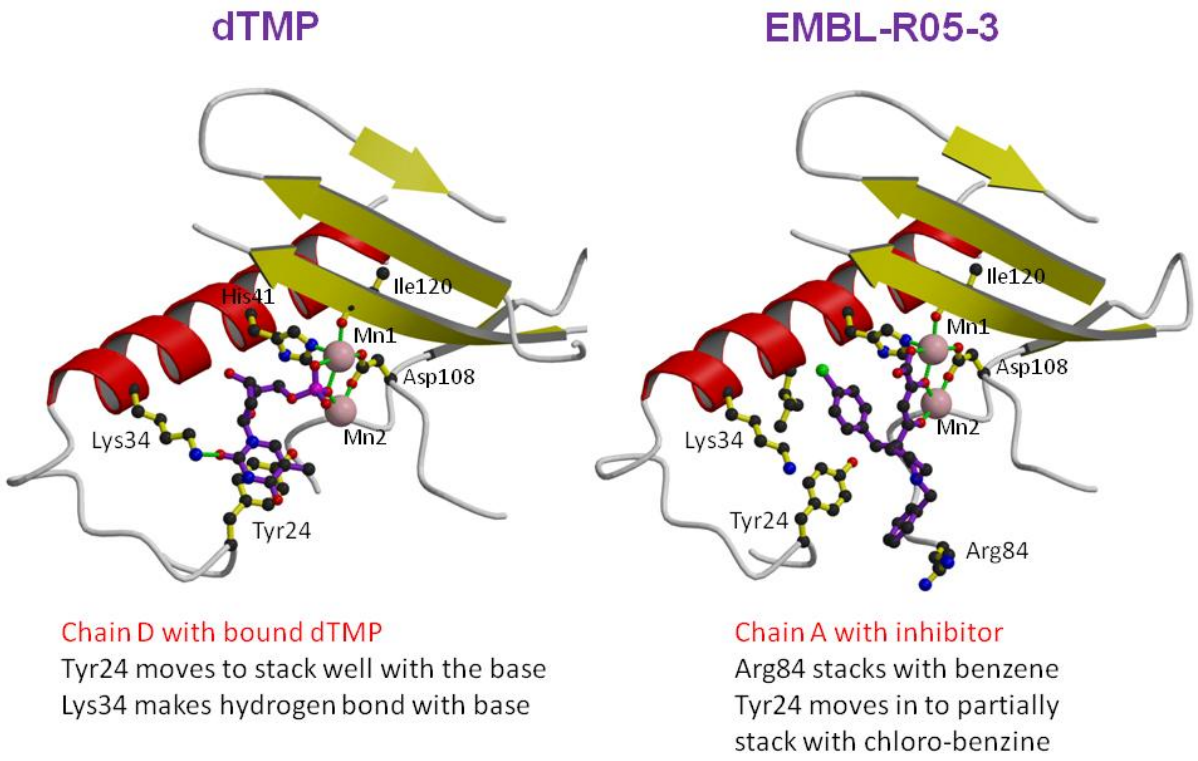


Figure 15. Comparison of dTMP/rUMP binding to that of EMBL R05-3.

To examine this further we superposed various complexes of restriction enzymes with bound substrate or product dsDNA complexes. Superposing for instance, the active sites of PA-Nter with EcoRV restriction enzyme closely overlaps the metal binding center (see (Dias et al, 2009). As seen in Figure 16, the bound rUMP most closely mimics the position of the post cleavage nucleotide. The fact that no other nucleotide could be observed in a bound form suggests that there may be some specificity in this post-cleavage position. Further work is required on the structures of PA-Nter substrate or product complexes to confirm this.

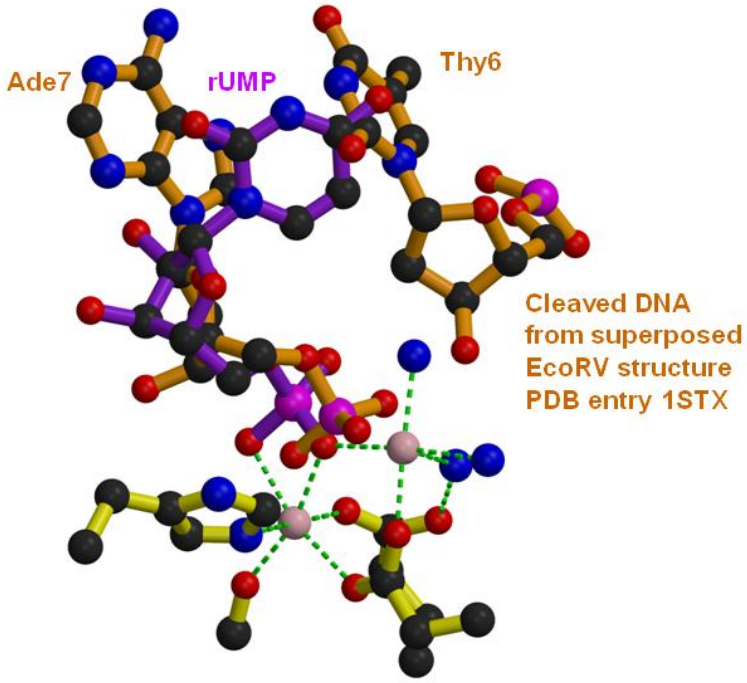


Figure 16. **Active site of H5N1 PA with bound rUMP with superposed product complex from EcoRV (pdb entry 1STX).** Active site residues, magnesium ion and water molecules are as in Figure 10 with bound rUMP in purple. The position of the two DNA bases either side of the cleavage site in the EcoRV product complex is shown with orange sticks.

Methods

Cloning expression and purification of H1N1 PA endonuclease domain.

The DNA coding for PA-N-ter (residues 1-198) from A/California/04/2009-H1N1 was synthesized and sub-cloned in the expression vector pESPRIT002 (EMBL) by GeneArt, (Regensburg, Germany). The sequence was designed to contain an MGSGMA polypeptide linker between the tobacco etch virus (TEV) cleavage site at the N-terminus to obtain 100% cleavage by TEV protease, as used previously (Dias et al, 2009).

Sequence of A/California/04/2009-H1N1 used:

mgsgma(1)MEDFVRQCFNPMIVELAEKAMKEYGEDPKIETNKFAAICTHLEVCFMYSD
FHFIDERGESIIVESGDPNALLKHRFEIIEGRDRIMAWTVVNSICNTTGVEKPKFLPDL
YDYKENRFIEIGVTRREVHIYYLEKANKIKSEKTHIHIFSFTGEEMATKADYTLDEESR
ARIKTRLFTIRQEMASRSLWDSFRQSERGE -198

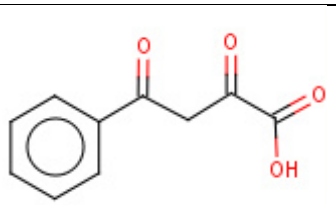
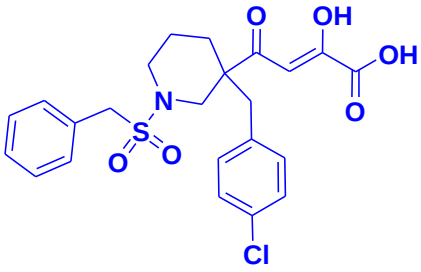
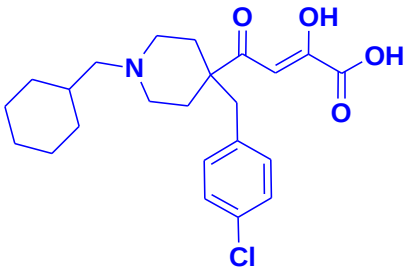
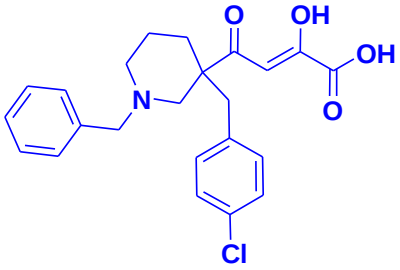
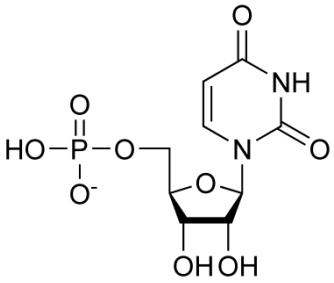
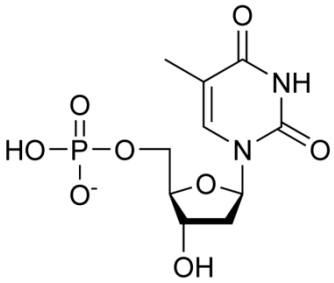
To potentially improve crystallisation properties, a deletion of part of the flexible loop (52-73) was engineered by site directed mutagenesis. For this a PCR amplification of the whole vector containing the wild type gene was performed using two primers flanking the mutation site, one of them phosphorylated, and TurboPfu polymerase (Stratagene). Subsequently template vector was digested with DpnI (New England Biolabs) and the mutated vector was re-ligated. In the mutant amino acid sequence 52-64 (HFIDERGESIIVE) was replaced by a single glycine.

Wild type and mutant plasmids were transformed to *E. coli* BL21(DE3) (Stratagene) and the protein was expressed in LB medium overnight at 20°C after induction at an OD 0.8-1.0 with 0.2 mM isopropyl-β-thiogalactopyranoside (IPTG). The protein was purified by an immobilized metal affinity column (IMAC). A second IMAC step was performed after cleavage by the His-tagged TEV protease, followed by gel filtration on a Superdex 75 column (GE Healthcare). Finally, the protein was concentrated to 10-15 mg.mL-1.

Compounds

Compounds used for co-crystallisation are given in Table 1. DPBA was purchased from Interchim and rUMP and rTMP from Sigma. Compounds EMBL-R05-01/02/03 (first described in (Tomassini et al, 1994)) were custom synthesised by Shanghai ChemPartner.

Table 1. Compounds used in this work.

Compound	CA Index Name	Formula	Chemical structure
DPBA	2,4-dioxo-4-phenylbutanoic acid	C10 H8 O4	
EMBL-R05-1 (Tomassini et al. Antimicrob Agents Chemother 1994, 38:2827-2837)	4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethylsulpho)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid	C23 H24 Cl N S O6	
EMBL-R05-2 (Tomassini et al. Antimicrob Agents Chemother 1994, 38:2827-2837)	4-[4-[(4-chlorophenyl)methyl]-1-(cyclohexylmethyl)-4-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid	C23 H30 Cl N O4	
EMBL-R05-3 (Tomassini et al. Antimicrob Agents Chemother 1994, 38:2827-2837)	4-[3-[(4-chlorophenyl)methyl]-1-(phenylmethyl)-3-piperidinyl]-2-hydroxy-4-oxo-2-butenoic acid	C23 H24 Cl N O4	
rUMP Ribo-uridine monophosphate	5'-Uridylic acid	C9 H13 N2 O9 P	
dTMP Thymidine monophosphate	5'-thymidylic acid	C10 H15 N2 O8 P	

Crystallization

Initial sitting drop screening was carried out at 20°C mixing 100 nL of protein solution (15 mg.mL-1) with 100 nL of reservoir solution using a Cartesian robot. Subsequently, larger crystals were obtained at 20°C by the hanging drop method mixing protein and reservoir solutions in a ratio of 1:1. The protein solution contained 10-15 mg.mL-1 of PA-N-ter in 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM $MnCl_2$, 2 mM $MgCl_2$. The refined reservoir compositions for native crystals and co-crystallization with different ligands are listed in Table 2. Native crystals and those co-crystallized with DBPA and EMBL R05-3 were flash frozen in liquid nitrogen after cryo protection in their reservoir solution containing 25% glycerol. Co-crystals with dTMP and rUMP were frozen in their reservoir solution containing 20% glycerol and 10 mM dTMP or rUMP, respectively. Structures of R05-2 and R05-1 were obtained by soaking co-crystals of PA-N-ter and dTMP for 2h with reservoir solution containing the inhibitor followed by cryo protection in reservoir solution containing 20% glycerol and the inhibitor.

Table 2. Summary of crystallisation conditions and crystallographic parameters for various compounds

Protein	Ligand (final concentration)	Crystallisation reservoir condition	Space group and unit cell parameters	Resolution Refinement R-factor/R-free
PA H1N1 1-198	Native	1.6 M sodium formate 0.1 M HEPES pH7 5% Glycerol	*C*2 4 molecules/ASU 263.63 66.24 66.32 90.00 95.98 90.00	2.1 Å 0.238 /0.278
PA H1N1 1-198	DBPA	1M LiCl2 0.1M HEPES 7 20% PEG6000	C2 4 molecules/ASU 260.87 65.75 66.03 90.00 95.76 90.00	2.3 Å 0.219/0.286
PA H1N1 1-198	EMBL-R05-3 (1.5 mM)	2.0M ammonium sulphate 0.1M BisTris pH5.5	*P* 2_1 2_1 2_1 4 molecules/ASU 54.57 122.54 129.78 90.00 90.00 90.00	2.5 Å 0.205 /0.276
PA H1N1 1-198 Δ52-64:Gly	EMBL-R05-1 (1.5 mM) soaked into crystals initially grown	25-30% PEG4K 0.1M Tris pH8.5 0.2M NaCl	*P* $6_2$22 (180) 1 molecule/ASU a=b= 75.06 c=120.05 Å	1.9 Å 0.201/0.251

	with dTMP)			
PA H1N1 1-198	EMBL-R05-2 (1.5 mM, soaked into crystals initially grown with rUMP)	0.1 M ammonium sulphate 0.1 M Bis-Tris pH5.5 25 % (w/v) PEG 3350	*P* 2_1 2_1 2_1 4 molecules/ASU 56.59 120.81 128.20 90.00 90.00 90.00	2.07 Å 0.205 /0.259
PA H1N1 1-198	rUMP (5.0 mM)	0.1 M ammonium sulphate 0.1 M Bis-Tris pH5.5 25 % (w/v) PEG 3350	*P* 2_1 2_1 2_1 4 molecules/ASU 54.94 120.11 128.05 90.00 90.00 90.00	2.05 Å 0.207 /0.249
PA H1N1 1-198	dTMP	25%PEG3350 Bis-tris pH5.5 0.2M ammonium sulphate	*P* 2_1 2_1 2_1 4 molecules/ASU 54.89 121.24 129.02 90.00 90.00 90.00	1.87 Å 0.196/0.237

Crystal structure determination.

Diffraction data were collected at 100 K on various beamlines at the European Synchrotron Radiation Facility (Table 3). Datasets were integrated with XDS (Kabsch, 1993) and scaled with XSCALE. Subsequent data analysis was performed with the CCP4i programme suite. The initial H1N1 structure was solved with molecular replacement with PHASER (Read, 2001) using the previously determined H3N2 PA N-ter structure (Dias et al, 2009). Subsequent co-crystal structures were determined with PHASER using the H1N1 structure. Refinement was carried out with REFMAC (Murshudov, 1997) and model building with COOT or O. In the *C*2 and $P2_12_12_1$ crystal forms there are four molecules per asymmetric unit. However due to structural variations between the molecules due to plasticity (in particular the 53-73 region) and the generally good resolution, NCS restraints were not applied.

Anomalous scattering from manganese was readily observed for the $P2_12_12_1$ and $P6_222$ crystal forms at the X-ray energies of normal data collection. This was not the case for the *C*2 crystal form (native and DPDA data), probably due to the lower symmetry. A separate dataset of the native C2 crystals was collected at a wavelength of 1.55 Å (see main text).

The sequence alignment in Figure 1 was drawn with ESPript (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi)(Gouet et al, 1999). Structure figures were

drawn with MOLSCRIPT (Kraulis, 1991) or BOBSCRIPT (Esnouf, 1999) and rendered with Raster3D (Merritt & Bacon, 1997).

Acknowledgements.

The authors would like to thank members of the ESRF-EMBL Joint Structural Biology Group for access to ESRF beamlines and the EMBL High-throughput crystallisation platform for crystallisation facilities and the UVHCI influenza group for discussions.

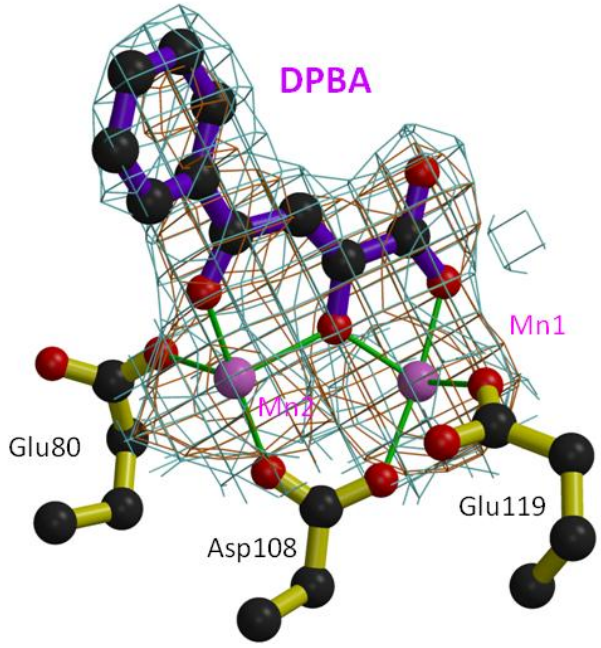
Table 3. Crystallographic data.

Ligand	Native	DPBA	R05-1	R05-2
Protein	H1N1 PA 1-198	H1N1 PA 1-198	H1N1 PA 1-198 Δ52-64:G	H1N1 PA 1-198
Beamline	ESRF ID14-4	ESRF ID29	ESRF ID29	ESRF ID23-1
Wavelength (Å)	0.9395	0.9474	0.9724	0.9724
Detector	ADSC Quantum Q315r	ADSC Quantum Q315r	ADSC Quantum Q315r	ADSC Quantum Q315r
Space Group	*C*2	*C*2	$P6_222$	$P2_12_12_1$
Cell dimensions (Å)	a= 263.63 b=66.24 c=66.32 β=95.98	a= 260.87 b=65.75 c=66.03 β=95.76	a=b= 75.06 c=120.05	a= 56.59 b=120.81 c=128.20
Solvent content (%)				
Molecules in ASU	4	4	1	4
Resolution range (last shell) (Å)	50-2.1 (2.2-2.1)	50-2.3 (2.4-2.3)	50-1.9 (1.95-1.90)	50-2.2 (2.3-2.2)
Completeness (last shell) (%)	99.0 (97.9)	97.3 (95.2)	99.9 (100)	99.6 (99.8)
R-sym (last shell) (%)	9.3 (65.8)	9.5 (39.7)	4.7 (53.7)	6.8 (51.8)
I/sigmaI	10.63 (2.2)	7.45 (2.16)	22.19 (2.95)	14.68 (2.44)
No. of unique reflections used in refinement (free)	63353 (2643)	46575 (1899)		42983 (2307)
R-factor (last shell)	0.238 (0.328)	0.219 (0.265)	0.201 (0.317)	0.199 (0.234)
R-free (last shell)	0.278 (0.353)	0.286 (0.383)	0.251 (0.355)	0.254 (0.348)
TLS/NCS used	No, No	No, No	No, -	No, No
No.of atoms : Protein Ligand Ions Solvent	 6572 - 4 x (Mg^{2+}, Mn^{2+}) 184	 6547 4 x DPDA 4 x (Mn^{2+}, Mn^{2+}) 272	 1592 1 x R05-1 1 x (Mn^{2+}, Mn^{2+}) 107	 6854 4 x R05-3 4 x (Mn^{2+}, Mn^{2+}) 303
Mean B-value ($Å^2$)	27.9	20.8	30.7	36.2
Ramachandran plot (Molprobity) Favoured regions				
Allowed regions				
Disallowed regions				
RMS deviations from ideal				

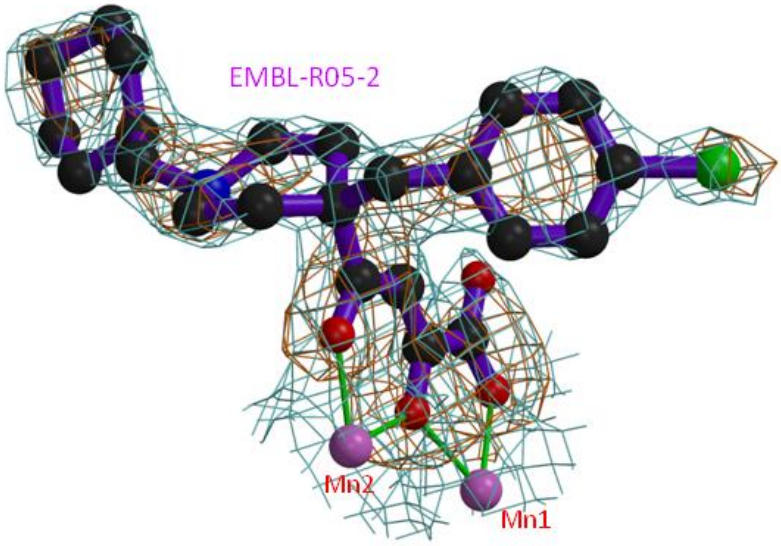
values Bond distances (Å)	0.023	0.021	0.022	0.022
Angles	1.80	1.72	1.843	1.898

Table 3. Crystallographic data (continued).

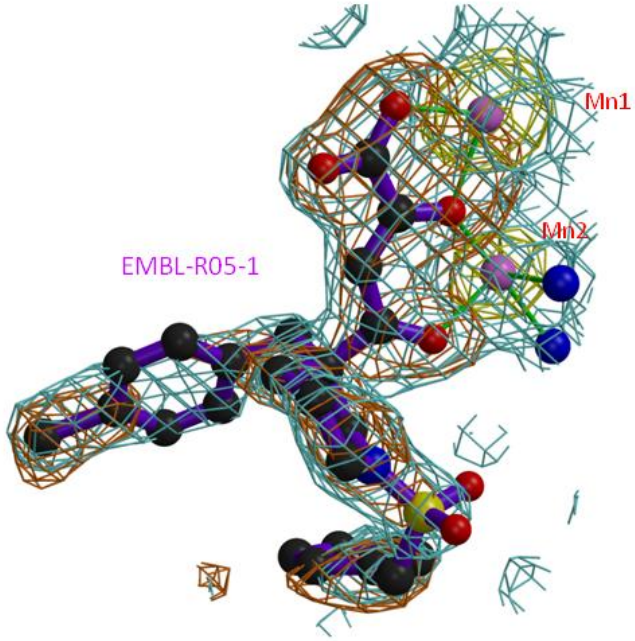
Ligand	R05-3	dTMP	rUMP
Protein	H1N1 PA 1-198	H1N1 PA 1-198	H1N1 PA 1-198
Beamline	ESRF ID23-2	ESRF ID14-4	ESRF ID23-1
Wavelength (Å)	0.8726	0.9765	0.9724
Detector	MAR225	ADSC Quantum Q315r	ADSC Quantum Q315r
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)	a= 54.57 b=122.54 c=129.78	a= 54.89 b=121.24 c=129.02	a= 54.96 b=120.20 c=128.07
Solvent content (%)			
Molecules in ASU	4	4	4
Resolution Range (last shell) (Å)	50-2.6 (2.7-2.6)	50-1.87 (1.92-1.87)	50-2.3 (2.10-2.05)
Completeness (last shell) (%)	99.8 (99.5)	99.8 (99.9)	99.5 (99.7)
R-sym (last shell)(%)	15.5 (81.2)	6.0 (86.9)	4.8 (50.7)
I/sigmaI	9.50 (2.14)	16.83 (2.22)	11.36 (1.62)
No. of unique reflections used in refinement (free)	28997 (1545)	68174 (3655)	51176 (2750)
R-factor (last shell)	0.205 (0.268)	0.196 (0.304)	0.207 (0.265)
R-free (last shell)	0.276 (0.378)	0.237 (0.333)	0.249 (0.340)
TLS, NCS used	Yes, No	Yes, No	No, No
No.of atoms : Protein Ligand Ions Solvent	 6523 4 x R05-3 4 x (Mn^{2+}, Mn^{2+}) 120 + 14 x SO_4	 6677 4 x dTMP 4 x (Mn^{2+}, Mn^{2+}) 266	 6688 4 x UMP 4 x (Mn^{2+}, Mn^{2+}) 272
Mean B-value ($Å^2$)	29.1	39.6	32.5
Ramachandran plot (Molprobity) Favoured regions			
Allowed regions			
Disallowed regions			
RMS deviations from ideal values Bond distances (Å)	 0.016	 0.024	 0.022
Angles	1.513	1.89	1.78



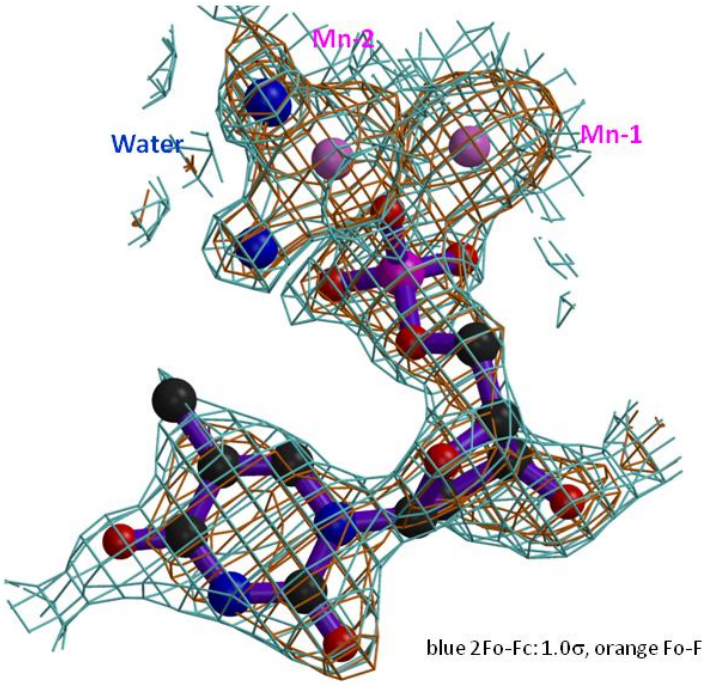
Supplementary Figure 1: Electron density for bound DPBA**.**



Supplementary Figure 2: Unbiased electron density for EMBL-R05-02 bound in H1N1 PA endonuclease. Blue contour: final 2Fo-Fc electron density at 1.0σ, orange contour Fo-Fc unbiased difference map: 2.7σ. Manganese ions are pink spheres and the ion co-ordination is shown with green lines.

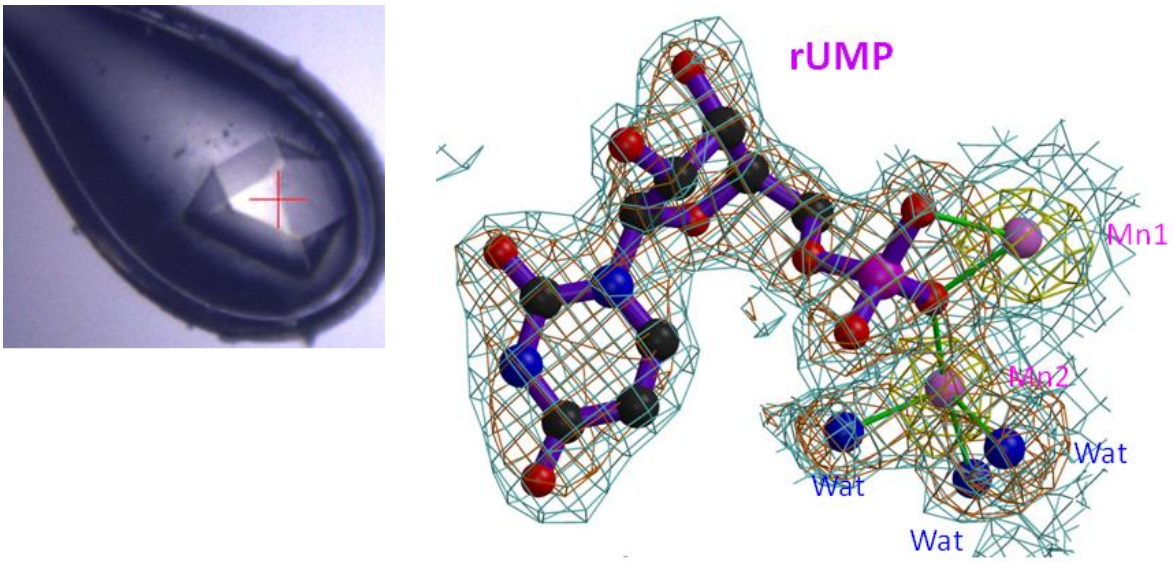


Supplementary Figure 3. **Unbiased electron density for EMBL-R05-01 bound in H1N1 PA endonuclease.** Blue contour: final 2Fo-Fc electron density at 1.0σ, orange contour Fo-Fc unbiased difference map: 2.8σ, yellow contour anomalous density at 5.0 σ. Manganese ions are pink spheres and co-ordinating water molecule blue spheres.



Supplementary Figure 4. Unbiased electron density for dTMP bound in H1N1 PA endonuclease. Blue contour: final 2Fo-Fc electron density at 1.0σ, orange contour Fo-Fc

unbiased difference map: 2.8σ. Manganese ions are pink spheres and co-ordinating water molecule blue spheres.



Supplementary Figure 5. Left: Frozen crystal of H1N1 PA-Nter co-crystallised with rUMP in the $P2_12_12_1$ space-group**. Right.** Blue contour: final 2Fo-Fc electron density at 1.1σ, orange contour Fo-Fc unbiased difference map: 2.8σ, yellow contour anomalous density at 4.0σ. Manganese ions are pink spheres and co-ordinating water molecule blue spheres and the ion co-ordination is shown with green lines.

References

Cianci C CT, Meanwell N, Putz H, Hagen M, Colonno RJ, Krystal M. (1996) Identification of N-hydroxamic acid and N-hydroxyimide compounds that inhibit influenza virus polymerase. *Antiviral Chemistry and Chemotherapy* 7(6): 353-360

**Crepin T, Dias A, Cusack S, Ruigrok RW, Palencia A, Swale C (2010) Mutational and metal binding analysis of the endonuclease domain
of the influenza virus polymerase PA subunit. *JVirology***

Das K, Aramini JM, Ma LC, Krug RM, Arnold E Structures of influenza A proteins and insights into antiviral drug targets. *Nature structural & molecular biology* 17(5): 530-538

Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 458(7240): 914-918

Doan L, Handa B, Roberts NA, Klumpp K (1999) Metal ion catalysis of RNA cleavage by the influenza virus endonuclease. *Biochemistry* 38(17): 5612-5619

Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical biochemistry* 357(2): 289-298

Esnouf RM (1999) Further additions to Molscript version 1.4, including reading and contouring of electron density maps. *Acta Crystallogr* 55: 938-940

Gong J, Fang H, Li M, Liu Y, Yang K, Liu Y, Xu W (2009) Potential Targets and Their Relevant Inhibitors in Anti-Influenza Fields. *Current medicinal chemistry*

Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics (Oxford, England)* 15(4): 305-308

Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ, Cusack S (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature structural & molecular biology* 15(5): 500-506

Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464(7286): 232-236

Hastings JC, Selnick H, Wolanski B, Tomassini JE (1996) Anti-influenza virus activities of 4-substituted 2,4-dioxobutanoic acid inhibitors. *Antimicrobial agents and chemotherapy* 40(5): 1304-1307

Iwai Y, Takahashi H, Hatakeyama D, Motoshima K, Ishikawa M, Sugita K, Hashimoto Y, Harada Y, Itamura S, Odagiri T, Tashiro M, Sei Y, Yamaguchi K, Kuzuhara T Anti-influenza activity of phenethylphenylphthalimide analogs derived from thalidomide. *Bioorg Med Chem* 18(14): 5379-5390

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. ***J Appl Cryst*** **26: 795-800**

Kraulis PJ (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. ***J Appl Cryst*** **24: 946-950**

Krug RM, Aramini JM (2009) Emerging antiviral targets for influenza A virus. ***Trends in pharmacological sciences*** **30(6): 269-277**

Kuzuhara T, Iwai Y, Takahashi H, Hatakeyama D, Echigo N (2009) Green tea catechins inhibit the endonuclease activity of influenza A virus RNA polymerase. ***PLoS Curr Influenza*****: RRN1052**

Merritt EA, Bacon DJ (1997) Raster3D Photorealistic molecular graphics. ***Methods in enzymology*** **277: 505-524**

Murshudov GN (1997) Refinement of macromolecular structures by the maximum-likelihood method. ***Acta crystallographica*** **53(Pt 3): 240-255**

Nakazawa M, Kadowaki SE, Watanabe I, Kadowaki Y, Takei M, Fukuda H (2008) PA subunit of RNA polymerase as a promising target for anti-influenza virus agents. ***Antiviral research*** **78(3): 194-201**

Parkes KE, Ermert P, Fassler J, Ives J, Martin JA, Merrett JH, Obrecht D, Williams G, Klumpp K (2003) Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors. ***Journal of medicinal chemistry*** **46(7): 1153-1164**

Plotch SJ, Bouloy M, Ulmanen I, Krug RM (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. ***Cell*** **23(3): 847-858**

Poon LL, Pritlove DC, Fodor E, Brownlee GG (1999) Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. ***Journal of virology*** **73(4): 3473-3476**

Read RJ (2001) Pushing the boundaries of molecular replacement with maximum likelihood. ***Acta crystallographica*** **57(Pt 10): 1373-1382**

Reguera J, Weber F, Cusack S (2010) Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. ***PLoS pathogens*** **6(9)**

Ruigrok RW, Crepin T, Hart DJ, Cusack S (2010) Towards an atomic resolution understanding of the influenza virus replication machinery. ***Curr Opin Struct Biol*** **20(1): 104-113**

Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M, Fiore F, Gardelli C, Gonzalez Paz O, Hazuda DJ, Jones P, Kinzel O, Laufer R, Monteagudo E, Muraglia E, Nizi E, Orvieto F, Pace P, Pescatore G, Scarpelli R, Stillmock K, Witmer MV, Rowley M (2008) Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. ***Journal of medicinal chemistry*** **51(18): 5843-5855**

Tomassini J, Selnick H, Davies ME, Armstrong ME, Baldwin J, Bourgeois M, Hastings J, Hazuda D, Lewis J, McClements W, et al. (1994) Inhibition of cap (m7GpppXm)-dependent endonuclease of

influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. ***Antimicrobial agents and chemotherapy*** **38(12): 2827-2837**

Tomassini JE, Davies ME, Hastings JC, Lingham R, Mojena M, Raghoobar SL, Singh SB, Tkacz JS, Goetz MA (1996) A novel antiviral agent which inhibits the endonuclease of influenza viruses. ***Antimicrobial agents and chemotherapy*** **40(5): 1189-1193**

Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T, Fodor E, Rao Z, Liu Y (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. ***Nature*** **458(7240): 909-913**

Zhao C, Lou Z, Guo Y, Ma M, Chen Y, Liang S, Zhang L, Chen S, Li X, Liu Y, Bartlam M, Rao Z (2009) Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate binding site inside the catalytic center. ***Journal of virology*** **83(18): 9024-9030**

Chapter 5

Methods

Summary of methods chapter

The basis for structural studies of biologic processes is to have pure, homogeneous and stable macromolecules to work with. In a typical work flow for protein production, a gene sequence coding for a particular protein is cloned into a plasmid vector. The vector is introduced in an expression host, i. e. bacterial cells, insect cells or mammalian cells. The host is then stimulated to actually produce the artificial gene product. Usually the protein is tagged with a specific short peptide sequence. This tag has a certain affinity for a specific molecule or metal which facilitates the separation of the protein of interest from the rest of cellular proteins. If after this step the protein is not sufficiently pure, further chromatographic purification steps (ion exchange chromatography, size exclusion chromatography, etc.) can be necessary. In the case of RNA and DNA purifications specific properties such as charge and size of the molecules are used to separate the desired molecule from side products of the reaction. The pure products can be subjected to biochemical and structural studies.

Because of the wavelength limit of visible light, protein molecules are too small to be resolved in a normal light microscope. Thus, methods like light scattering, electron microscopy (EM), nuclear magnetic resonance (NMR) or crystallography have to be used to observe them. Structural biology methods used in this thesis include scattering techniques that analyze molecules and molecular shapes in solution and crystallography to determine macromolecular structures at atomic resolution. Nevertheless, structural techniques always require complementary functional experiments. Vice versa, functional studies benefit from structural knowledge.

Résumé du chapître "Méthodes"

La base des études structurales des processus biologiques est la disponibilité de manière pure, homogène et stable des macromolécules étudiées. Dans le cas typique de la production d'une protéine, la séquence du gène codant est clonée dans un vecteur d'expression qui est lui-même introduit dans une lignée de cellules d'expression (i. e. les cellules bactériennes, les cellules d'insectes ou cellules de mammifères). Cet hôte est ensuite stimulé afin de produire la correspondante au gène cible. La protéine d'intérêt est séparée du reste des protéines cellulaires à l'aide d'une courte séquence d'acides aminés par affinité. Si, après cette étape, la protéine n'est pas encore assez pure et homogène, d'autres étapes de purification peuvent être nécessaires : chromatographie par échange d'ions, chromatographie d'exclusion stérique, etc... Pour la purification des nucléotides (ARN et des ADN), les propriétés spécifiques de la molécule (comme la charge et la taille) sont mises à profit pour séparer la molécule désirée des produits secondaires réactionnels. Après plusieurs étapes de purification (chromatographie, électrophorèse préparative) les molécules pures peuvent alors être soumises à des études biochimiques.

La taille des molécules biologiques, comme les protéines ou les acides nucléiques, est trop petite pour qu'elles soient visualisées par un microscope optique. Ainsi, des méthodes comme la diffusion de la lumière, la microscopie électronique (EM), la résonance magnétique nucléaire (RMN) ou des méthodes cristallographiques doivent être utilisées pour les observer et les caractériser. Les méthodes de biologie structurale utilisées dans ce manuscrit sont principalement des techniques biophysique basées sur la diffusion de lumière (SAXS et SANS) et la cristallographie au rayon X. Néanmoins, les techniques structurales exigent nécessairement des expériences fonctionnelles (biochimie) complémentaires. Réciproquement, les études fonctionnelles bénéficient de la connaissance de la structure des macromolécules étudiées lors des études structurales.

5.1 Genetic engineering methods

Gene sequences for human RIG-I and CARDIF were kindly provided by Daniel Kolakofsky, Geneva, the template for human MDA5 and measles leader RNA by Denis Gerlier, Lyon and human Linear ubiquitin clones by David Komander, Cambridge. The clone for production of influenza short pan handle (shPH) RNA was kindly provided by Thibaut Crepin, as well as an aliquot of readily purified signal recognition particle (SRP) *Alu* domain RNA of *Pyrococcus horikoshii* and a *Candida alibicans* $tRNA^{Asn}$. All other sequences were purchased as genes optimized for expression in the referring host (bacterial or insect cell expression) at GeneArt, Regensburg, Germany.

Vectors used in this work were pETM-11, pETM-13, pETM-30, pETM-40, pETM-44, pRSET, pCDF-Dual and pESPRIT002 for bacterial expression and pFastBac Htb and pFastBac Dual for insect cell expression. The following *E.coli* cell lines were used: TOP10 (Invitrogen), MACH1 (Invitrogen), BL21star(DE3) (Invitrogen), ROSETTA2-(DE3) (Merck), DH10Bac (Invitrogen).

For amplification of fragments up to 4 kbp, primers were designed to include a priming sequence of at least 23 base pairs (bp) corresponding to the amplified gene, a cutting site for a specific restriction enzyme and if necessary, purification tags and start or stop codons. The theoretic melting temperature of the priming sequence was adjusted to 55 °C. The primers were purchased from Invitrogen.

Fragments were amplified by polymerase chain reaction (PCR) in a reaction mix (50 μl) containing 0.5 μl of each primer (100 μM stock), 1 μl dNTP mix (10 mM each, Fermentas), 1x Pfu buffer (Stratagene), 20-200 ng DNA template and 1 μl Pfu polymerase (stock 2.5 U/ml, core facility, EMBL Heidelberg). The reaction was performed in a thermal cycler (Biometra) with the program for a touch down PCR described in table 5.1. After amplification of the DNA fragment, the PCR mix was analyzed by agarose gel elec-

trophoresis (0.5-1.5% agarose, in TAE (for 50 l: 242 g Tris, 100 ml 0.5 M EDTA pH 8.0), running buffer 1x TAE, loading buffer: for 100 ml 10x buffer: 250 mg bromophenol blue, 250 mg xylene cyanol, 33 ml 150 mM Tris pH 7.6, 60 ml glycerol). The band corresponding to the amplified fragment was excised from the gel and purified using a gel extraction kit (Quiagen). Subsequently, the fragment and the target vector were digested with a restriction enzyme, following the instruction of the enzyme provider (New England Biolabs). The digested PCR fragments were then purified with the PCR purification kit (Qiagen) and the vector was purified from an agarose gel after electrophoresis. Vector and insert were ligated following the instructions of the ligation kit "Mighty mix" (TaKaRa) and ligated 30 min to overnight at 16 °C. A 5 μl aliquot of the reaction mix was transformed to chemically competent TOP10 cells. When the vector required a streptomycine resistance, competent MACH1 cells were used. For transformation, cells were thawed 5 min on ice, then the plasmid was added and cells were incubated for further 30 min. After a heat shock of 45 s at 42 °C, cells remained for further 2 min on ice before the addition 250 μl of LB medium and incubation for 1 h at 37 °C. Cells were spread on LB agar plates containing the suitable selective antibiotic and grown overnight. Single colonies were picked, grown in 5 ml cultures overnight at 37 °C. Plasmid was isolated using a Miniprep kit (Promega or Qiagen) and its sequence was verified by sequencing reactions.

For gene mutation, site-directed mutagenesis was performed, using the Stratagene QuickChange. For this, gene and vector are amplified together in a circular manner. For mutagenesis of 1-3 consequent bases a primer was designed to contain the mutation site in the middle and 15-25 flanking bases at each side. The T_m of mutation primers was supposed to be >75 °C. The second primer contained the reverse complementary sequence. For longer deletions, primers were designed upstream and downstream of the deletion site (T_m=55 °C). One of the primers was phosphorylated on the 5'-end. The reaction mix (25 μl) for these reactions contained 1x Pfu buffer (Stratagen), 5-50 ng DNA tem-

Table 5.1: *Cycles for PCR (Pfu Polymerase)*

number of cycles	temperature	time
	98 °C	1 min
3x	98 °C	30 s
	63 °C	30 s
	72 °C	30 s per 500 bp
3x	98 °C	30 s
	61 °C	30 s
	72 °C	30 s per 500 bp
3x	98 °C	30 s
	59 °C	30 s
	72 °C	30 s per 500 bp
5x	98 °C	30 s
	57 °C	30 s
	72 °C	30 s per 500 bp
18x	98 °C	30 s
	55 °C	30 s
	72 °C	30 s per 500 bp
	98 °C	10 min
	4 °C	hold

plate, 125 ng of each primer, 1 μl dNTP mix (10 mM each, Fermentas) and 1.25 U Turbo Polymerase (Invitrogen). The touch down thermo cycler program for this site directed mutagenesis is described in table 5.2. After the plasmid-amlification, 20 U DpnI (New England Biolabs) were added to eliminate the template plasmid. Ligation and transformation were carried out as described previously.

Very short constructs up to about 100 bp, for example templates for transcription of short RNAs, were produced by assembly of synthesized oligonuceotides. For this, oligonucleotides containing sequence and restriction sites and their overlapping reverse complements up to the length of 60 bp were annealed by heating to 95 °C and slow cooling down to room temperature. Then the fragments were directly ligated to a cleaved but not dephosphorylated vector and transformed in TOP10 cells as described previously.

Table 5.2: *Cycles for mutagenesis PCR (Turbo Polymerase)*

number of cycles	temperature	time
	95°C	1 min
3x	95°C	30 s
	62°C	30 s
	68°C	30 s per 500 bp (insert + vector)
3x	95°C	30 s
	60°C	30 s
	68°C	30 s per 500 bp (insert + vector)
5x	95°C	30 s
	58°C	30 s
	68°C	30 s per 500 bp (insert + vector)
5x	95°C	30 s
	56°C	30 s
	68°C	30 s per 500 bp (insert + vector)
15x	95°C	30 s
	54°C	30 s
	68°C	30 s per 500 bp (insert + vector)
	98°C	10 min
	4°C	hold

5.2 RNA production and purification

Preparation of DNA template RNA was produced by T7 transcription. To prepare the transcription, the DNA template was linearized by enymatic restriction. The sequence contained from 5' to 3': the T7 promoter sequence (5'TAATACGACTCACTATAGGG3'), the desired sequence preferably staring with a G or GG, and an enzymatic restriction site. 1 mg of DNA was incubated with 2000 Units of restriction enzyme in its required buffer (New England Biolabs) at 37°C overnight. The linearized DNA was extracted by two steps of phenol/chlorophorm extraction and one step of chlorophorm extraction. The DNA was precipitated by addition of 3 volumes of 100% ethanol and 1/3 volume of sodium or potassium acetate and incubated for 30 min at -80°C. The precipitated DNA was pelleted, washed with 80% ethanol, dried and dissolved in water.

T7 transcription The T7 transcription was performed in 40 mM Tris/HCl pH 8.0, 5 mM DTT, 0.01% Triton X100, 30 mM $MgCl_2$, 4 mM of each NTP (100 mM stocks, buffered at pH 7.5), 8% PEG 8000, 0.1 mg/ml linear template DNA and 40 μg/ml T7 RNA polymerase (10 mg/ml stock). The mix was incubated for 4-5 hours at 37°C. Af-

ter, precipitated magnesuium pyrophosphate was resolubilized by addition of EDTA to a final concentration of 50 mM. Then, urea was added to a final concentration of 4 M. The RNA was purified either by a PAGE or alternatively by size exclusion chromatography. Purification by PAGE was found to be more suitable for our relatively small RNAs. [88].

RNA purification by PAGE The transcription mix containing 4 M urea and the dyes bromophenol blue, xylencyanol and orange G was loaded on a polyacrylamide gel (1x TBE buffer (for 10 l: 108g Tris, 55 g boric acid, 40 ml 0.5 M Na_2EDTA pH 8.0), 8 M urea, 8-20% acrylamide (40% stock; 19:1BisAA, 1% ammonium persulfate (APS), 0.5% TEMED). Electrophoresis was carried out for 5-8 hours at a maximal power of 65-85 Watts. Bands were visualized by cutting a small stripe of the gel which was stained in methylene blue. The desired band was excised and inserted in an gel electro elution chamber (Elutrap). The RNA was eluted overnight at 120 V in a buffer 1x TBE and 2 M urea. Subsequently, the RNA was desalted on a PD10 column (GE Healthcare) or dialyzed against water. In a final step, the RNA was concentrated to the desired volumne in an speedvac.

RNA purification size exclusion chromatography The transcription mix was supplemented with urea to a final concentration of 4 M and loaded on an S75(16/60) column, which was previously equilibrated in 4 M urea. Elution fractions were analyzed by PAGE. Selected fractions were pooled and the urea was removed by dialysis against water. After, the RNA was precipitated by the addition of ethanol and sodium or potassium acetate and incubated at -20 °C for 20 min. The pelleted RNA was dissolved in water, desalted on a PD10 column (GE Healthcare) and concentrated using a speedvac or an Amicon concentrator.

5.3 Protein expression and purification

5.3.1 Recombinant protein expression in *E. coli*

For recombinant protein expreesion in *E.coli*, an expression plasmid was transformed to a suitable expression cell line like BL21*(DE3) or Rosetta2(DE3). Clones from LB-agar plates were inoculated in liquid LB medium with the appropriate antibiotics. The cultures were upscaled from 2 ml to 50 ml to 1 L at 37 °C on a shaker. The cells were grown to an A_{600} 0.8 to 1 and induced by 0.2 mM IPTG (isopropyl-β-D-thiogalactopyranoside). After expression overnight at 20 °C on a shaker, the cells were harvested by centrifugation at 3500 g for 15 minutes at 4 °C . Cell pellets were frozen and stored at -80 °C or directly processed.

Rich medium for auto-induction

For expression in auto-inducible (AI) medium the following solutions were prepared: 1 M $MgSO_4$ (24.65 g $MgSO_4$ in 100 ml H_2O), 50x5052 (for 100 ml: 25 g glycerol, 2.5g glucose, 10 g α-lactose), 20xNPS (for 100 ml: 6.6 g $(NH_4)_2SO_4$. 13.6 g KH_2PO_4, 14.2 g Na_2HPO_4, pH 6.75) and ZY (10 g tryptone, 5 g yeast extract, 925 ml H_2O). The solutions were sterilized separately and mixed before inoculation of bacteria; for 1 L: 1ml $MgSO_4$, 20 ml 50x5052, 50 ml 20xNPS, antibiotics and ZY to make 1 L. A small volume of preculture (in LB medium) was inoculated and the culture was grown at 37 °C on a shaker. After the 2 h the temperature was lowered to 35 °C and the culter was incubated on a shaker overnight [135].

5.3.2 Recombinant protein expression in baculo virus system

For cloning and expression in insect cells, both the protocol for Bac-to-Bac baculo virus expression systems (Invitrogen) and the protocol by Fitzgerald *et al.* [40] were followed. For expression in a baculo virus system, it is necessary to clone the gene fragment into a suitable vector, like pFastBac Htb or pFastBac Dual (Invitrogen). After sequence ver-

ification, the vector is transposed to a bacmid. For this, competent DH10 bac cells are thawed 5 min on ice prior to the addition of 5 ng of plasmid. After a further 30 min of incubation on ice, cells were heat shocked at 42 °C for 45 s and after 2 more minutes on ice 900 μl of LB medium was added to the cells. Cultures were incubated overnight at 37 °C, before plating undiluted culture and cultures diluted 1:10 and 1:100 on agar plates containing 50 μg/ml kanamycin, 7 μg/ml gentamycin, 10 μg/ml tetracyclin and 100 μg/ml Bluo-gal and 40 μg/ml IPTG for blue white selection. Due to destruction of a LacZ gene upon successful transposition, positive colonies stay white. White colonies and were selected and inoculated in 2 ml liquid LB cultures. After overnight culture, cells were lysed and genomic bacterial DNA was precipitated with the three solutions of a plasmid purification kit (Promega or Qiagen). After repeated centrifugation to generate a very clear solution, plasmid DNA was precipitated with isopropanol and the cell pellet was washed with 80% ethanol. After short drying the bacmid DNA was resolved in 20 μl of water.

Sf21 cells (Invitrogen) from a shaking culture were transfered to 6 well plates to a density of 0.5 to 0.8 10^6 cells per well and incubated at 26 °C for 15-30 min for attachment. In the meantime, the 20 μl of bacmid was mixed with 200 μl of Sf900 medium and 10 μl of Fugene HD (Roche) and incubated for 15-30 min. Subsequently, the bacmid medium mixture was distributed to two wells of attached cells and the total medium volume per well was adjusted to 3 ml. After 2-3 days the V_0 virus is released to the medium and can be harvested by aspiration. For amplification of the virus, 50 ml of Sf21 cells at a density of 0.5 10^6 cells/ml were infected with 3 ml of V_0 virus. Cells were split daily to 0.5 10^6 cells/ml until cell growth stops, then they are cultured for 2 more days before clearing the medium containing the V_1 virus from cell debris by centrifugation (maximum 800 g for 20 min) and filtration.

The V_1 virus as used to infect HiFive cells for protein expression. For this, HiFive cells at a density of 1 10^6 cells/ml were infected with 2 %(v/v) V_1 virus. Depending on

protein expression, cells were harvested after 60 h to 3 days by centrifugation (maximum 800 g for 20 min). Cell pellets were frozen and stored at -80 °C or directly processed.

Seleno-methionine incorporation in insect-cell expressed protein

Se-Met incorporation in the baculovirus expressed protein was achieved following a protocol of Cronin and co-workers [27]. Hi5 cells were adapted to ESF 921 (Expression systems, Woodland, CA) by maintaining a culture in the new medium until they divided normally. Cells were infected at a density of 0.5 10^6 cells/ml with 1% (v/v) virus. After a certain time period the medium was exchage to methionine free ESF 921. After further incubation to deplete residual methionine, L(+)-Selenomethionine (Acros) was added. The incubation times and the amount of seleno-methionine has to be individually optimized for each virus. Cells were harvested after 2-3 days.

5.3.3 Protein purification

Buffers used for in typical purification of His-tagged protein are listed in table 5.3. For the purification of PA N-ter and its mutants 2 mM $MgCl_2$ and 2 mM $MnCl_2$ was added to all buffers. Cell pellets were thawed if necessary and resolved in buffer A with protease inhibitors (Complete, Roche). Cells were disrupted by sonication or two passages through a microfluidizer. The cell debris was separated from soluble cell components by centrifugation at 40 000 g for 30 min at 4 °C. In a standard case, proteins were tagged by a 6x histidine tag (His-tag) and the cleared lysate was loaded on a nickel affinity resin (Ni-NTA, Quiagen, binding capacity $\sim$ 10mg/ml). Due to the chemical properties of a multi-histidine tag the protein is bound to metal ions coordinated on the column. The interaction between histidines and ions can be interrupted by elevated concentrations of imidazole in the washing buffers. The column was washed with 10 column volumes (CV) of buffer B with high salt and elevated imidazole concentration to remove residual nucleic acids. The high salt wash was omitted in purification of protein complexes. After, the column was washed with 10 CV of buffer A and the protein was eluted with buffer E.

Table 5.3: *Protein purification buffers*

	A	B	E	D	E2	G
HEPES pH 7.5	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM
NaCl	150 mM	1 M	150 mM	150 mM	150 mM	150 mM
2-Mercaptoethanol	5 mM	5 mM	5 mM	1 mM	5 mM	5 mM
Glycerol	2%	2%	2%	2%	2%	
Imidazole	20mM	40 mM	300 mM		40 mM	

For removal of the His-tag, 1mM DTT and 1 mg TEV protease per 100 mg of protein were added and incubated overnight at 4 °C. TEV protease cleaves the amino acid sequence ENLYFQG inserted before the protein sequence to remove the the His-tag. After, the protein solution was filled into an dialysis bag (Spectrum Laboratories) with the molecular weight cut off of 6-8 kDa and the buffer was exchanged to buffer D (see 5.3) by dialysis to remove DTT and imidazole. After a minimum of 6 h of dialysis, the protein solution was loaded on a 2nd Nickel column and the flow through was collected. In this step the His-tagged TEV protease and the cleaved His-tags remain bound to the column, but also contaminants binding unspecifically to the 1st nickel column bound Ni-resin stay bound at low imidazole concentrations. The column was washed with buffer E2 until no protein was eluted anymore. Depending on individual protein properties, the protein of interest was in the flow through or in the E2 elution fraction.

Protein containing fractions were concentrated by centrifugation in Amicon concentrators with a suitable molecular weight cut-off. As a final polishing step, the protein was purified on a size exclusion column. The stationary phase of such a column consists of porous beads with a well-defined range of pore sizes. Small globular molecules require longer time to run through the column whereas large elongated molecules pass the column without lingering. No more than 5 mg of protein and no more than a volume of 0.5 ml were loaded on a S75(10/300) or S200(10/300) which was pre-equilibrated with buffer G.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Throughout purification, loading, washing and eluted fractions were analyzed on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (separating gel: 7-15% acrylamide (30%/0.8% w/v), 0.375 M Tris pH 8.8, 0.1% sodium dodecyl sulfate (SDS), 1% ammonium persulfate (APS), 0.5% TEMED; stacking gel: 2.5% acrylamide (30%/-0.8% w/v), 0.25 M Tris pH 6.8, 0.1% sodium dodecyl sulfate (SDS), 1% ammonium persulfate (APS), 0.5% TEMED; running buffer for 10 L: 10 g SDS, 30.3 g Tris, 144.1 g glycin; 2x loading buffer for 100 ml 1.5 Tris pH 6.8, 6 ml 20% SDS, 30 ml glycerol, 15 ml β-mercaptoethanol and 1.8 mg bromophenol blue). Protein concentration was checked with a Bradford assay (Biorad) or using a Nanodrop photospectrometer taking into account the theoretical absorption coefficient calculated by the software PROTPARAM [144].

Other tags

In the case of a protein fused to a purification tag different from a multi His-tag, buffers did not contain imidazole and a different strategy was used for elution. Maltose binding protein (MBP) tagged proteins were loaded onto an amylose resin (New England Biolabs; binding capacity ∼3 mg/ml). Bound protein was eluted from the column in elution buffer containing 10 mM maltose. For glutathione S-transferase (GST) tagged proteins the affinity column consisted of gluthathione linked to the column matrix (Amersham Biosciences; binding capacity ∼10 mg/ml resin). The GST-protein is eluted with a buffer containing 10 mM reduced glutathione. Strep tag II tagged proteins were bound to a Strep-Tactin column (IBA; binding capacity ∼50 nmol/ml resin) and eluted in a buffer containing 2.5 mM desthiobiotin.

5.3.4 Protein methylation

To facilitate crystallization of highly soluble proteins, it is possible to modify the protein surface characteristics by methylation of surface lysins. A 1 M formaldehyde stock and

a dimethylamne boran complex (ABC) stock solution was prepared (weighing 60 mg of ABC powder in 1 ml of buffer). In a first step, protein solution concentrated to maximal 0.5 mg/ml in a non-Tris buffer was supplemented by 20 μl ABC solution per 1 ml of protein solution and 400 μl 1 M formaldehyde per 1 ml of protein solution, gently mixed and incubated at 4 °C. After 2 h this procedure is repeated. After a further 2 h, additional 10 μl ABC solution per 1 ml of protein solution were added and the reaction was incubated at 4 °C overnight. Aggregated protein was removed by centrifugation before stopping the reaction by a buffer exchange to Tris on a desalting column or dialysis.

5.3.5 Biochemical analysis

Western Blotting

In a western blot, proteins are specifically detected by means of antibodies. An SDS-PAGE was performed and the blot was transferred from the gel by onto a PVDF (Polyvinylidenfluorid) sheet. Before blotting, the PVDF sheet was soaked in 100% ethanol and in western blotting buffer (for 1 l: 1 g SDS, 3.03 g Tris, 14.41 g glycin, 10% ethanol). A sandwich was built (from negative to positive electrode), containing 4 layers of Whatman paper, the gel, the PVDF sheet and 4 more layers of Whatman paper. The samples were blotted for 1 hour at 100 V at 4 °C. After the transfer, the PVDF sheet was saturated in 2.5% milk powder, 0.1% tween-20 in 1x PBS to reduce unspecific binding of the antibody added in the next step. For anti- anti-his-tag antibody conjugated to horseradish peroxidase (HRP) in a dilution 1:2000 in the milk buffer from above was incubated for 1 h. For other blots the a primary antibodz was incubated for 1 h followed by 2 washing steps with the milk buffer of 15 min before the secondary antibody was applied for 1 h. After antibody incubation, two wash steps of 15 minutes were performed before detection. For direct color-detection on the membrane, SIGMAFAST 3,3-Diaminobenzidine tablets were dissolved in 1 ml of water following the protocol and the membrane was stained. The staining reaction was stopped by additon of water. For photodectetion, ECL Plus Wester Blotting Reagents (Amersham Biosciences) were used and photo sensitive

film was exposed and devolopped following the protocol of the kit.

Blotting for N-terminal sequencing

For N-terminal sequencing, the protein was transferred from an SDS gel onto a PVDF membrane in a 3-(cyclohexylamino)-(propanesulfonic acid) (CAPS) buffer blotting technique. Before blotting, the PVDF sheet was shortly soaked in 100% methanol and after equilibrated in transfer buffer (10 mM CAPS pH 11, 10% methanol). In the same time the SDS gel was washed in water and equilibrated in the same transfer buffer. The protein was blotted as descibed for the western blot and the protein was detected on the membrane with 0.1% Coomassie blue R-250 in 40% methanol and 1% acetic acid for one minute. Excess colorant was removed by repeated washes in 50% methanol. Finally the gel was rinsed with water. Bands of interest were excised from the dried membrane and analyzed by N-terminal sequencing at the Institute Biologie Structurale, Grenoble.

ATPase assay

For the ATPase assay 50 mM Tris/HCL pH 7, 5 mM DTT, 2 mM $MgCl_2$, 0.5 mM ATP mix, 0.1 mM RNA and protein in different concentrations, varying from 15 μM to 200 pM were incubated at 37 °C for 15 min. The ATP mix consisted of a mixture of ATP:ATPγ in a ratio of 15:1. Control reactions without RNA were set up. After 15 min the reaction was was stopped by the addition of 4 μl formic acid and 3 μl of each reaction were spotted on a thin layer chromatography (TLC) plate (PEIF Cellulose) and migrated in 0.5 M LiCl and 1 M formic acid in a migration chamber for 30 min. Free phosphates migrate faster than the ATP or ADP and the radioactive signal of the freed phosphate can be counted as a measure for ATP hydrolysis.

Thermal shift assay

In the thermal shift assay (or Thermofluor), the protein sample is slowly heated up and heat-induced denaturation is detected by a hydrophobic fluoroprobe. In solution this

probe is quenched. Binding to the exposed hydrophobic core of the unfolding protein activates the fluorophor. The fluorescence emission is plotted as a function of temperature. The melting temperature (T_m) is calculated to be the temperature at the inflection point of the emission curve. The higher the T_m, the more stable the protein and the higher is the probability of the protein to crystallize [38]. For the thermal shift assay 1-100 μM of protein (5 μM in standard case) were mixed with with 1 μl of Sybro Orange and various buffer solutions or additives in a real time (RT) PCR plate. In the RT PCR machine, the plate was heated up from 25-°Cto 73 °C by an increment of 1 °C per minute and the emission of the fluoroprobe was measured. The dye was excited at 490 nm and the emission light was recorded at 575 nm.

Native polyacrylamide gel electrophoresis (PAGE)

On native PAGE, proteins or protein complexes migrate dependent on their charge, shape and weight. The protein or protein RNA sample was mixed with 4x native loading buffer (for 50ml: 25 ml 0.5 M Tris/HCl pH 6.8, 20 ml glycerol, 0.5 mg bromphenol blue), and applied on an native acrylamide gel (7-15% acrylamide (30%/0.8%w/v)), 0.375 M Tris pH 8.8, 1% ammonium persulfate (APS), 0.5%TEMED). The native PAGE was run 3h at 4 °C in native running buffer (for 2 l: 12 g Tris, 57.6g glycine). Native gels were stained with ethidium bromide for RNA and with InstantBlue stain (Invitrogen) for protein. Alternatively native TBE gels were used. For this a native TBE gel was cast (1x TBE, 8-20% acrylamide 40% stock; 19:1BisAA, 1% ammonium persulfate (APS), 0.5% TEMED) and samples were loaded with DNA loading buffer. TBE gels were run for 2 h at 200 V at 4 °C and stained with InstantBlue for protein or ethidium bromide for RNA.

Limited proteolysis

Limited proteolysis can determine fragments of a protein or protein complex which are resistant to proteolytic digestion. Those fragments are compact globular domains (since inaccessible for the protease) and are likely to crystallize. For limited proteolysis, protein

solution was mixed with trypsin in a ratio of 1:500 to 1:1000. Samples were taken at several time points and the reaction was stopped by mixing with SDS loading buffer and heating to 95 °C for 5 min. For large scale purifications of trypsinated protein the reaction was stopped by the addition of protease inhibitor (Complete, Roche) and the mixture was immediately purified by size exclusion chromatography to remove residual trypsin.

In vitro Ubiquitination assay

Purified constructs of RIG-I (20 μM) and TRIM25 (5 μM) were incubated in a reaction buffer (50 mM Tris-HCl, 2 mM dithiothreitol, 5 mM $MgCl_2$ and 4 mM ATP) with 25 μM ubiquitin (Sigma), 0.1 μM human recombinant E1 (BIOMOL) and 2.5 *mu*M human recombinant UbcH5a (BIOMOL) at 32°C for 2 h.

Surface plasmon resonance (SPR) assay or Biorad assay

Surface plasmon resonance enables label-free analysis of biomolecular interactions. The technique is based on a phenomenon that occurs when polarized light, under the condition of total internal reflection, strikes an electrically conducting gold layer at the interface between media of different refractive index. In our case this is the glass of a sensor surface (high refractive index) and a buffer (low refractive index). The signal is a measure for the mass concentration adsorbed onto the sensor chip suface. Analyte and ligand association and dissociation from the chip can be observed and rate constants as well as equilibrium constants can be calculated.

Surface plasmon resonance (SPR) experiments were performed on a BIAcoreX instrument using a CM5 sensor chip (BR-1000-14). First, the chip was activated by EDC (1-ethyl-3-(3-dimethzlaminoprppyl)-carbodiimide) and NHS (N-hydroxysuccinimide). Then, the protein was bound to the chip in a buffer of a pH lying in between pH 3 (the pH of the chip) and the PI of the corresponding protein. In our experiment, TRIM25 PRYSPRY (434-630) was immobilized in 10 mM sodium acetate pH 4. After, the surface was

deactivated by the application of ethanolamine. During immobilization the signal was monitored to reach stable baselines. Then the analyte, RIG-I CARDs(1-226) and RIG-I CARDs(1-226)T55I, respectively, was injected in a concentration range of 50 to 0.625 μM in 20 mM HEPES pH 7.5 and 100 mM NaCl. Association and dissociation signal was measured. In between the runs of different analyte concentrations, the chip was washed with NaCl to remove the ligand. Curve fitting was performed with the Biacore software using a 1:1 binding model. The quality of the curve fit was estimated with the help of the chi^2 value, which should be $\leq$2. Final K_D values were averaged from the curves of the concentrations from 40 μM to 5 μM.

5.4 Scattering techniques

Scattering techniques like multi angle laser light scattering (MALLS), small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) enable structural analysis of macromolecular particles in homogeneous solutions.

5.4.1 Multi Angle Laser Light Scattering (MALLS)

With the method of Multi Angle Laser Light Scattering MALLS it is possible to determine the exact molecular weight of particles in solution. Combined with size exclusion chromatography (SEC-MALLS) it is used to determine the exact mass of macromolecues eluting from a size exclusion chromatography column. In biological applications this is applied to determine the stoichiometry of macromolecular complexes as well as their homogeneity. In a typical experiment, a light scattering detector and a concentration detector like an refractive index detector are connected in line after the gel filtration column. For SEC-MALLS measurements 40 μl of a solution containing 2-4 mg/ml protein and ligands in the desired ratio were injected in an S200(10/300) column, connected to a MALLS detector (DAWN-EOS Wyatt technology) and a refractive index detector (RI2000b Schambeck). Data were analyzed with the ASTRA V software [159].

5.4.2 Small Angle Scattering (SAS)

Small angle scattering techniques analyze molecular shape, quaternary and tertiary structure of molecules in solution. The resulting low-resolution envelopes that can be combined with known high-resolution structures. In contrast to electron microscopy, the method can be used for smaller macromolecules and is therefore complementary. Small angle neutron scattering (SANS) provides additional information for macromolecular complexes that are made of several types of molecules such as proteins, nucleic acids (DNA or RNA) or lipids. Contrast variation experiments obtained by exchanging the solvent for deuterated or partially deuterated solvent. Due to the different neutron scattering properties of nucleotides and protein, it is possible to choose the H_2O:D_2O ratio of the buffer such that on or other component does not contribute to the scattering. Contrast matching is achieved at about 43% D_2O for protein and 73% D_2O for RNA. Even single components of protein complexes can be expressed deuterated. Models obtained by small angle X-ray scattering (SAXS) have slightly higher resolution then SANS data. For the experiment the sample solution is exposed to the radiation and the scattered intensity is detected dependent on the scattering angle. The Scattering of the buffer only is subtracted as background. The random positions and orientations of particles result in an isotropic intensity distribution which, for mono disperse non-interacting particles, is proportional to the scattering from a single particle averaged over all orientations. Since distinct shapes (globilar, disc-like, elongated) have characteristic scattering curves, a particle fitting to the result scattering curve is simulated. Several fits are combined to a model of the molecular envelope of the molecule [41].

For SAXS and SANS experiments, protein solution were prepared in different concentrations ranging from 1 mg/ml to 10 mg/ml. For measurements of protein-RNA complexes with SANS, the protein buffer was exchanged from a water-based buffer to a the same buffer containing 73% D_2O or 100% D_2O on a PD-10 desalting column (Amersham). Data was analyzed by means of the ATSAS software suite [6]; raw data were

processed with PRIMUS, the distance distribution functions were analyzed with GNOM, *ab initio* models were built with DAMMIN or DAMMIF. The programme DAMAVER was used for model averaging, SUPCOMB to superpose models and the program CRYSOL to convert pdb model data into SAS curves.

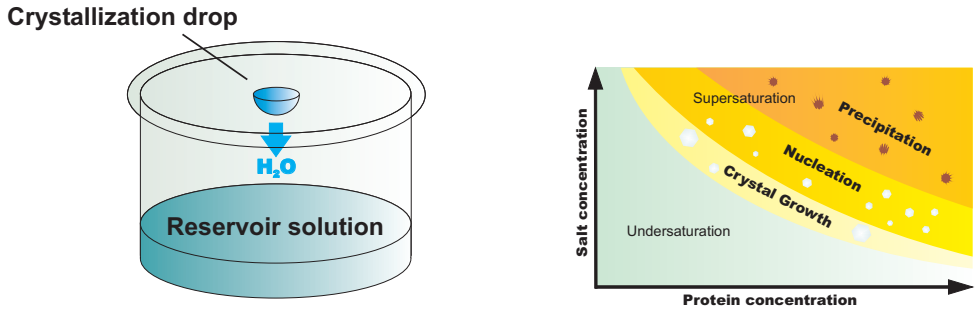
5.5 X-ray crystallography

Electron microscopy, SAXS and SANS reveal relatively low resolution molecular structures, while nuclear resonance spectroscopy (NMR) and X-ray crystallography are able to determine high resolution atomic three dimensional atomic structures. After successful expression and purification of a protein, the next major hurdle is to grow crystals of the molecule. Once available crystals are then irradiated with a focused and intense monochromatic X-ray beam and the diffraction pattern is detected. From these pattern, the electron density of the molecules in the crystal can be calculated and a model of the molecule can be built into the density.

5.5.1 Crystallization by vapor diffusion

For crystallization by vapor diffusion, a hanging or sitting drop of protein-precipitant mixture is enclosed with a reservoir of precipitant solution. As the concentration of precipitant is lower in the droplet than in the reservoir, water molecules are transferred through the vapor phase to the bottom solution until equilibrium is reached. Thereby, the protein concentration in the droplet rises slowly. When saturation is reached, nucleation occurs in rare cases in the protein drop and some protein molecules may organize into a highly ordered crystal structure. In the case of undersaturation, the droplets stay clear, in the case of too high protein and precipitant concentration, the protein precipitates. (Figure 5.1).

By default, 6 x 96 crystallization conditions were screened with the htx robot at EMBL Grenoble in in the sitting drop set-up. Screens included Hampton: crystal screen I and II, crystal screen lite, PEG/Ion screen, MembFac, Natrix, QuickScreen, ammo-



(a) (b)

Figure 5.1: *Crystallization by vapor diffusion. (**A**) A typical hanging drop crystallization set-up. One volume of reservoir solution is added to a distinct volume of protein mixture and sealed together with reservoir solution. Water molecules diffuse from the droplet to the reservoir until equilibrium is reached. (**B**) A schematic phase diagram. At very high salt and protein concentrations the protein in the droplet will precipitate. In the unsaturated zone, the molecules stay in solution. For nucleation and crystal growth, the solution should generally be only slightly saturated.*

nium sulfate grid screen, sodium malonate grid screen, PEG 6K grid screen, PEG/LiCl grid screen, MPD grid screen, Index screen; htx-made screens: screen Mme, sodium formate screen). Typically, protein concentrations rangeing from 5 mg/ml to 20 mg/ml were tested. Those drops had a size of 100 nl protein solution plus 100 nl reservoir solution. The reservoir contained 100 μl of reservoir solution. In the case of a hit in such a screen, subsequent grid screens varying slightly precipitant concentrations and pH of the hit-condition were performed to obtain single crystals of suitable size. Hanging drops were set up onto siliconized cover glasses or in 24-well sitting drop plates in 1:1 (protein to precipitant) ratio. The wells were sealed with the cover glasses by means of high viscous silicone paste or with transparent film.

5.5.2 Cryo protection of crystals

During X-ray exposure, crystals suffer from radiation damage as free radicals form. To minimize this effect, the crystals are typically cooled down to about 100 K during the experiment. Since the formation of ice crystals would destroy the protein crystal, this is prevented by a cryo-protection before flash-freezing the crystals in liquid nitrogen. As

cryo-protectants, low molecular weight PEGs, glycerol or ethyle glycol are commonly used. The crystals were briefly soaked in the cryo protectant solution containing the reservoir conditions and a certain concentration and 25-30% glycerol or ethylene glycol. In the case of co-crystallization of a protein with small molecules (inhibitors, nucleotides) those were added to the cryo protectant solution to prevent soaking out of this molecules.

5.5.3 Data collection

Data were collected at ESRF beamlines, Grenoble, France. A data collection strategy was designed by analyzing several diffraction images with the program MOSFLM [75].

5.5.4 Data processing

Data was indexed and integrated with XDS [64]. Indexing determines the relative orientation of the unit cell and the crystal symmetry of the crystal from collected data. Subsequently, the data were integrated by measuring the difference in intensity between the spots and the local background. Then, the data were merged and scaled with XSCALE [64]. Structure factor amplitudes were calculated from merged intensities by the program TRUNCATE implemented in CCP4 [24]. Another challenge in X-ray crystallography is the phase problem. In a diffraction experiment, the scattered waves from the crystal planes contain all information about the structure. The amplitudes are measured in the experiment, but the phase information is lost. To solve the phase problem, heavy atoms can be introduced into the protein molecule, as the more electrons an atom possesses, the more it scatters X-rays. Small differences in reflection intensity between native and derivative can be used to first locate the heavy atoms and then perform phasing (isomorphous replacement). Anomalous scattering occures when atoms not only scatter photons but also absorb energy due to energy transition of electrons. The required energy is characteristic for the chemical element and nature of transition. At wavelengths close to a absorption edge the atomic scattering factor includes an anomalous component and the phase of the emitted photon is altered, which can be used to obtain phases. The SIRAS

(Single isomorphous replacement with anomalous scattering) phasing was carried out with SHARP [32]. The second possibility to find phases is molecular replacement [117]. In molecular replacement a known structure with a certain structural similarity (at least 25% sequence identity as a rule of thumb) is used to obtain the phases. Molecular replacement was performed with PHASER [133, 114], implemented in CCP4 [24]. The Matthews coefficient was calculated to determine the number of molecules in the asymmetric unit by estimating the solvent content [97].

5.5.5 Refinement and structure analysis

As a better model gives back better phases, several rounds of refinement have to be performed starting with the initial model. For this, the structure was inspected and improved, using the molecular graphics program COOT [37]. After a cycle of building, a refinement program was run to minimize the R-factor, which is a measure for congruency of the modeled structure factors $|F_{calc}|$ with the observed ones $|F_{obs}|$:

$$R = \frac{\Sigma||F_{obs}|-|F_{calc}||}{\Sigma|F_{obs}|}$$

The programs REFMAC [24] and ARP/WARP [73] were used for refinement. For the calculation of R_{free}, a test set of 5% of the reflections were excluded from refinement to circumvent model bias [15].

$$R_{free} = \frac{\Sigma_{test\ set}||F_{obs}|-|F_{calc}||}{\Sigma_{test\ set}|F_{obs}|}$$

The program ARP/WARP [73] was used to build water molecules. A number of further programs and software were used for analysis of the structure. PYMOL [111] was used to created figures, CLUSTALX and CLUSTALW [74] were used for sequence alignments. Alignments were drawn using ESPRIPT [51]. For structure predictions the NS and for geometry check of the structure the MOLPROBITY server was used [18].

Acknowledgements

This thesis has been an important part of my life for four years, it took more than a hundred pages to summarize and its realization would not have been possible without the support of many people, that I want to thank now.

I want to express my gratitude to Stephen Cusack for giving me the opportunity to work as a part of his group. Thank you for the freedom and the responsibility you gave me with the project. Thanks for brilliant ideas, sharp minded reflections, for your trust, and for keeping the scientific level always high above the everyday-lab-troubleshooting. I appreciate that you came up with a back-up plan at the right time!

While writing this, the day of the defense has not yet come, but I want to thank my jury members Winfried Weissenhorn, Daniel Kolakofsky and Denis Gerlier for revision of my work and for helpful suggestions. I would also like to thank my thesis advisory committee members Asifa Akthar, Ramesh Pillai and Winfried Weissenhorn for guidance throughout my PhD.

Limited by one's own knowledge, collaborations make it possible to push the limits and to get inspiration from a different point of view. For this I thank Phil Callow, Marc Jamin, Guy Schoehn, Christiane Schaffizel, Leandro Estrozi, Denis Gerlier, Jade Louber, Daniel Kolakowsky and Stephane Haussmann.

Furthermore, I am grateful to all past and present EMBL Grenoble group and team leaders, for making EMBL such an open-mided place always having an open door and open ear for constructive discussions and questions: Christoph Müller, Winfried Weissenhorn, Ramesh Pillai, Darren Hart, Daniel Panne, Imre Berger, Christiane Schaffizel and Andrew McCarthy. Thanks for their constant assistance to: the EMBL HTX team, the computer services team, the administrative and technical staff and the ESRF beamline staff.

This work would not have been possible without the steady interaction with the members of the Cusack lab. I thank Thomas Lunardi and Pavlo Holenya who contributed their work to this project. I want to thank Thibaut Crepin and Delphine Guilligay for helping me in the first months in the lab, Marcello Clerici for discussions, Julien Perard for support with RNA work and SAXS data and Jan Kadlec who guided me through structure solution. A big thanks to all past and present Cusack group mates: Cecile, Elena, Andrew, Thibaut, Delphine, Saurabh, Otto, Lars, Marcello, Andres, Jan, Juanito, Nikos, Tom, Michael, Chloe, Julien, Sergyi!

Thanks to the all-EMBL-predoc mates, especially predoc course 2006, to make scientific networking so much fun! And last but not least a big thanks to all the other EMBL Grenoble mates, to mention only some of them, Ricardo, Miriam, Victor, Jordi, Juliette, Manuela, Carmen and all the rest, for fun lunches, the ”international night” series, and one (or more) beers downtown.

Finally, I want to thank all my friends and my family for their support. Especially I thank my parents for constant encouragement. Why waste words here—you know it better anyway.

Dankeschön!

Bibliography

[1] Ablasser A, Bauernfeind F, Hartmann G, Latz E, Fitzgerald KA, Hornung V. RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. Nat Immunol. 2009 Oct.

[2] Andrejeva J, Childs KS, Young DF, Carlos TS, Stock N, Goodbourn S, Randall RE. The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter. Proc Natl Acad Sci U S A. 2004 Dec 7.

[3] Arias CF, Escalera-Zamudio M, Soto-Del Río Mde L, Cobián-Güemes AG, Isa P, López S. Molecular anatomy of 2009 influenza virus A (H1N1). Arch Med Res. 2009 Nov.

[4] Arimoto K, Takahashi H, Hishiki T, Konishi H, Fujita T, Shimotohno K. Negative regulation of the RIG-I signaling by the ubiquitin ligase RNF125. Proc Natl Acad Sci U S A. 2007 May 1.

[5] Ash and Roberts. Influenza: The State of Our Ignorance. Science. 21 April 2006.

[6] Atsas 2.3 program suite:
http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html

[7] Bao X, Kolli D, Liu T, Shan Y, Garofalo RP, Casola A. Human metapneumovirus small hydrophobic protein inhibits NF-kappaB transcriptional activity. J Virol. 2008 Aug.

[8] Barber MR, Aldridge JR Jr, Webster RG, Magor KE. Association of RIG-I with innate immunity of ducks to influenza. Proc Natl Acad Sci U S A. 2010 Mar 30.

[9] Baril M, Racine ME, Penin F, Lamarre D. MAVS dimer is a crucial signaling component of innate immunity and the target of hepatitis C virus NS3/4A protease. J Virol. 2009 Feb.

[10] Bartlam M, Lou Z, Liu Y, Rao Z. Polymerase structure and funciton. In: Influenza molecular virology. Wang Q and Tao YJ, editors. Caister Academic Press, 2010.

[11] Baum A, Sachidanandam R, García-Sastre A. Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. Proc Natl Acad Sci U S A. 2010 Sep 14.

[12] Beutler B. Microbe sensing, positive feedback loops, and the pathogenesis of inflammatory diseases. Immunol Rev. 2009 Jan.

[13] Bouchier-Hayes L, Martin SJ . CARD games in apoptosis and immunity. EMBO Rep 2002 3.

[14] Bouvier NM, Palese P. The biology of influenza viruses. Vaccine. 2008 Sep 12.

[15] Brünger AT. Assessment of phase accuracy by cross validation: the free R value. Methods and applications. Acta Crystallogr D Biol Crystallogr. 1993.

[16] Cao X. New DNA-sensing pathway feeds RIG-I with RNA. Nat Immunol. 2009 Oct.

[17] Cazenave C, Uhlenbeck OC. RNA template-directed RNA synthesis by T7 RNA polymerase. Proc Natl Acad Sci U S A. 1994 Jul 19.

[18] Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010 Jan.

[19] Chen YR and Clark AC. Equilibrium and kinetic folding of a á-helical Greek key protein domain: caspase recruitment domain (CARD) of RICK. Biochemistry 2003, 42.

[20] Chen YR and Clark AC (2004). Kinetic traps in the folding/unfolding of procaspase-1 CARD domain. Protein Sci. 2004.

[21] Childs K, Stock N, Ross C, Andrejeva J, Hilton L, Skinner M, Randall R, Goodbourn S. Mda-5, but not RIG-I, is a common target for paramyxovirus V proteins. Virology. 2007 Mar 1.

[22] Childs KS, Andrejeva J, Randall RE, Goodbourn S. Mechanism of mda-5 Inhibition by paramyxovirus V proteins. J Virol. 2009 Feb.

[23] Chiu YH, Macmillan JB, Chen ZJ. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. Cell. 2009 Aug 7.

[24] Collaborative Computing Project Number 4.The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1999 Sep.

[25] Cordin O, Banroques J, Tanner NK, Linder P. The DEAD-box protein family of RNA helicases.Gene 367, 17-37 (2006).

[26] Coussens NP, Mowers JC, McDonald C, Nunez G, Ramaswamy S. Crystal structure of the Nod1 caspase activation and recruitment domain. Biochem Biophys Res Commun. 2007 Feb 2.

[27] Cronin CN, Lim KB, Rogers J. Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. Protein Sci. 2007 Sep.

[28] Cui S, Eisenächer K, Kirchhofer A, Brzózka K, Lammens A, Lammens K, Fujita T, Conzelmann KK, Krug A, Hopfner KP. The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. Mol Cell. 2008 Feb 1.

[29] D'agostino PM, Amenta JJ, Reiss CS. IFN-beta-induced alteration of VSV protein phosphorylation in neuronal cells. Viral Immunol. 2009 Dec.

[30] Das K, Aramini JM, Ma LC, Krug RM, Arnold E. Structures of influenza A proteins and insights into antiviral drug targets. Nat Struct Mol Biol. 2010 May.

[31] De Falco F, Cainarca S, Andolfi G, Ferrentino R, Berti C, Rodríguez Criado G, Rittinger O, Dennis N, Odent S, Rastogi A, Liebelt J, Chitayat D, Winter R, Jawanda H, Ballabio A, Franco B, Meroni G. X-linked Opitz syndrome: novel mutations in the MID1 gene and redefinition of the clinical spectrum. Am J Med Genet A. 2003 Jul 15.

[32] De La Fortelle E, Bricogne G. Maximum-Likelihood Heavy-Atom Parameter Refinement for the Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. Methods in Enzymology 1997, 276.

[33] Deng L, Dai P, Parikh T, Cao H, Bhoj V, Sun Q, Chen Z, Merghoub T, Houghton A, Shuman S. Vaccinia virus subverts a MAVS-dependent innate immune response in keratinocytes through its dsRNA binding protein E3. J Virol. 2008 Aug 20.

[34] Desmyter J, Melnick JL, Rawls WE. Defectiveness of interferon production and of rubella virus interference in a line of African green monkey kidney cells (Vero). J Virol. 1968 Oct.

[35] Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature. 2009 Apr 16.

[36] Dixit E, Boulant S, Zhang Y, Lee AS, Odendall C, Shum B, Hacohen N, Chen ZJ, Whelan SP, Fransen M, Nibert ML, Superti-Furga G, Kagan JC. Peroxisomes are signaling platforms for antiviral innate immunity. Cell. 2010 May 14.

[37] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004 Dec.

[38] Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem. 2006 Oct 15.

[39] Fan L, Briese T, Lipkin WI. Z proteins of New World arenaviruses bind RIG-I and interfere with type I interferon induction. J Virol. 2010 Feb.

[40] Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I. Protein complex expression by using multigene baculoviral vectors. Nat Methods. 2006 Dec.

[41] Forstner M. J. Appl. Cryst. 2000. 33, 519-523. SAXS, SANS and X-ray crystallography as complementary methods in the study of biological form and function.

[42] Friedman CS, O'Donnell MA, Legarda-Addison D, Ng A, Cárdenas WB, Yount JS, Moran TM, Basler CF, Komuro A, Horvath CM, Xavier R, Ting AT. The tumour suppressor CYLD is a negative regulator of RIG-I-mediated antiviral response. EMBO Rep. 2008 Sep.

[43] Fujita T. A nonself RNA pattern: tri-p to panhandle. Immunity. 2009 Jul 17.

[44] Gack MU, Shin YC, Joo CH, Urano T, Liang C, Sun L, Takeuchi O, Akira S, Chen Z, Inoue S, Jung JU. TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. Nature. 2007 Apr 19.

[45] Gack MU, Albrecht RA, Urano T, Inn KS, Huang IC, Carnero E, Farzan M, Inoue S, Jung JU, García-Sastre A. Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. Cell Host Microbe. 2009 May 8.

[46] Gack MU, Kirchhofer A, Shin YC, Inn KS, Liang C, Cui S, Myong S, Ha T, Hopfner KP, Jung JU. Roles of RIG-I N-terminal tandem CARD and splice variant in TRIM25-mediated antiviral signal transduction. Proc Natl Acad Sci U S A. 2008 Oct 28.

[47] Gack MU, Nistal-Villán E, Inn KS, García-Sastre A, Jung JU. Phosphorylation-mediated negative regulation of RIG-I antiviral activity. J Virol. 2010 Apr.

[48] Gao D, Yang YK, Wang RP, Zhou X, Diao FC, Li MD, Zhai ZH, Jiang ZF, Chen DY. REUL is a novel E3 ubiquitin ligase and stimulator of retinoic-acid-inducible gene-I. PLoS One. 2009 Jun 1.

[49] Gitlin L, Barchet W, Gilfillan S, Cella M, Beutler B, Flavell RA, Diamond MS, Colonna M. Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. Proc Natl Acad Sci U S A. 2006 May 30.

[50] Grütter C, Briand C, Capitani G, Mittl PR, Papin S, Tschopp J, Grütter MG. Structure of the PRYSPRY-domain: implications for autoinflammatory diseases. FEBS Lett. 2006 Jan 9.

[51] Gouet P, Courcelle E, Stuart DI, Métoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999 Apr.

[52] Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ, Cusack S. The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat Struct Mol Biol. 2008 May.

[53] Guo Z, Chen LM, Zeng H, Gomez JA, Plowden J, Fujita T, Katz JM, Donis RO, Sambhara S. NS1 protein of influenza A virus inhibits the function of intracytoplasmic pathogen sensor, RIG-I. Am J Respir Cell Mol Biol. 2007 Mar.

[54] Habjan M, Andersson I, Klingström J, Schümann M, Martin A, Zimmermann P, Wagner V, Pichlmair A, Schneider U, Mühlberger E, Mirazimi A, Weber F. Processing of genome 5' termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction. PLoS ONE. 2008 Apr 30.

[55] Hausmann S, Marq JB, Tapparel C, Kolakofsky D, Garcin D. RIG-I and dsRNA-induced IFNbeta activation. PLoS One. 2008;3(12):e3965. Epub 2008 Dec 30.

[56] Hilbert M, Karow AR, Klostermeier D. The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. Biol Chem. 2009 Dec.

[57] Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucl. Acids Res. 38, W545-549, 2010.

[58] Hornung V, Ablasser A, Charrel-Dennis M, Bauernfeind F, Horvath G, Caffrey DR, Latz E, Fitzgerald KA. AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. Nature. 2009 Mar 26.

[59] Hornung V, Ellegast J, Kim S, Brzozka K, Jung A, Kato H, Poeck H, Akira S, Conzelmann KK, Schlee M, Endres S, Hartmann G. 5'-Triphosphate RNA is the ligand for RIG-I. Science. 2006 Nov 10.

[60] Ishikawa H, Barber GN. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature. 2008 Aug 24.

[61] James LC, Keeble AH, Khan Z, Rhodes DA, Trowsdale J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. Proc Natl Acad Sci U S A. 2007 Apr 10.

[62] Jankowsky E, Fairman ME. RNA helicases–one fold for many functions. Curr Opin Struct Biol. 2007 Jun.

[63] Jankowsky E, Putnam A. Duplex unwinding with DEAD-box proteins. Methods Mol Biol. 2010.

[64] Kabsch, WJ.Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J.Appl.Cryst. 26, 1993.

[65] Kato H, Takeuchi O, Mikamo-Satoh E, Hirai R, Kawai T, Matsushita K, Hiiragi A, Dermody TS, Fujita T, Akira S. Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. J Exp Med. 2008 Jul 7.

[66] Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, Uematsu S, Jung A, Kawai T, Ishii KJ, Yamaguchi O, Otsu K, Tsujimura T, Koh CS, Reis e Sousa C, Matsuura Y, Fujita T, Akira S. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. Nature. 2006 May 4.

[67] Kawai T, Akira S. Antiviral signaling through pattern recognition receptors. J Biochem (Tokyo). 2007 Feb.

[68] Kawai T, Akira S. SnapShot: Pattern-recognition receptors. Cell. 2007 Jun 1.

[69] Kawai T, Takahashi K, Sato S, Coban C, Kumar H, Kato H, Ishii KJ, Takeuchi O, Akira S. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. Nat Immunol. 2005 Oct.

[70] Keeble AH, Khan Z, Forster A, James LC. TRIM21 is an IgG receptor that is structurally, thermodynamically, and kinetically conserved. Proc Natl Acad Sci U S A. 2008 Apr 22.

[71] Kholod N, Vassilenko K, Shlyapnikov M, Ksenzenko V, Kisselev L. Preparation of active tRNA gene transcripts devoid of 3'-extended products and dimers. Nucleic Acids Res. 1998 May 15.

[72] Komander D, Reyes-Turcu F, Licchesi JD, Odenwaelder P, Wilkinson KD, Barford D. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. EMBO Rep. 2009 May.

[73] Langer GG, Cohen SX, Perrakis A, Lamzin VS. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature Protocols 3. 2008.

[74] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. ClustalW and ClustalX version 2. Bioinformatics 2007 23(21).

[75] Leslie, A.G.W. Recent changes to the MOSFLM package for processing film and image plate data Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, 1992, No. 26.

[76] Li T, Chen X, Garbutt KC, Zhou P, Zheng N. Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. Cell. 2006 Jan 13.

[77] Li X, Lu C, Stewart M, Xu H, Strong RK, Igumenova T, Li P. Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5. Arch Biochem Biophys. 2009 Aug 1.

[78] Li X, Ranjith-Kumar CT, Brooks MT, Dharmaiah S, Herr AB, Kao C, Li P. The RIG-I-like receptor LGP2 recognizes the termini of double-stranded RNA. J Biol Chem. 2009 May 15.

[79] Li XD, Sun L, Seth RB, Pineda G, Chen ZJ. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. Proc Natl Acad Sci U S A. 2005 Dec 6.

[80] Lin SC, Lo YC, Wu H. Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. Nature. 2010 Jun 17.

[81] Lindner P. Quick Guide: DEAD-box proteins. Curr Biol. 2000, 10.

[82] Ling Z, Tran KC, Teng MN.J Virol. Human respiratory syncytial virus nonstructural protein NS2 antagonizes the activation of beta interferon transcription by interacting with RIG-I. 2009 Apr.

[83] Lu C, Xu H, Ranjith-Kumar CT, Brooks MT, Hou TY, Hu F, Herr AB, Strong RK, Kao CC, Li P. The Structural Basis of 5' Triphosphate Double-Stranded RNA Recognition by RIG-I C-Terminal Domain. Structure. 2010 Jul 14.

[84] Manon F, Favier A, Nunez G, Simorre JP, Cusack S. Solution structure of NOD1 CARD and mutational analysis of its interaction with the CARD of downstream kinase RICK. J Mol Biol. 2007 Jan 5.

[85] Martinon F, Tschopp J (2004). Inflammatory caspases: linking an intracellular innate immune system to autoinflammatory diseases. Cell 117.

[86] Meylan E, Curran J, Hofmann K, Moradpour D, Binder M, Bartenschlager R, Tschopp J. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. Nature. 2005 Oct 20.

[87] Milam SL, Nicely NI, Feeney B, Mattos C, Clark AC. Rapid folding and unfolding of Apaf-1 CARD. J Mol Biol. 2007 May 25.

[88] Milligan JF, Uhlenbeck OC. Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol. 1989.

[89] Moon H, Choe J. Crystallization and preliminary crystallographic studies of human RIG-I in complex with double-stranded RNA. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2009 Jun 1.

[90] Moore CB, Bergstralh DT, Duncan JA, Lei Y, Morrison TE, Zimmermann AG, Accavitti-Loper MA, Madden VJ, Sun L, Ye Z, Lich JD, Heise MT, Chen Z, Ting JP. NLRX1 is a regulator of mitochondrial antiviral immunity. Nature. 2008 Jan 31.

[91] Murali A, Li X, Ranjith-Kumar CT, Bhardwaj K, Holzenburg A, Li P, Kao CC. Structure and function of LGP2, a DEX(D/H) helicase that regulates the innate immunity response. J Biol Chem. 2008 Jun 6.

[92] Myong S, Cui S, Cornish PV, Kirchhofer A, Gack MU, Jung JU, Hopfner KP, Ha T. Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. Science. 2009 Feb 20.

[93] Nishino T, Komori K, Tsuchiya D, Ishino Y, Morikawa K. Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing. Structure. 2005 Jan.

[94] Nisole S, Stoye JP, Saib A. TRIM family proteins: retroviral restriction and antiviral defence. Nat Rev Microbiol. 2005 Oct.

[95] Nistal-Villan E, Gack MU, Martínez-Delgado G, Maharaj NP, Inn KS, Yang H, Wang R, Aggarwal AK, Jung JU, Garcia-Sastre A. Negative role of RIG-I serine 8 phosphorylation in the regulation of interferon-beta production. J Biol Chem. 2010 Jun 25.

[96] Marq JB, Kolakofsky D, Garcin D. Unpaired 5' ppp-nucleotides, as found in arenavirus double-stranded RNA panhandles, are not recognized by RIG-I. J Biol Chem. 2010 Jun 11.

[97] Matthews BW. Solvent content of protein crystals. J Mol Biol. 1968.

[98] Minucci S, Maccarana M, Cioce M, De Luca P, Gelmetti V, Segalla S, Di Croce L, Giavara S, Matteucci C, Gobbi A, Bianchini A, Colombo E, Schiavoni I, Badaracco G, Hu X, Lazar MA, Landsberger N, Nervi C, Pelicci PG. Oligomerization of RAR and AML1 transcription factors as a novel mechanism of oncogenic activation. Mol Cell. 2000 May.

[99] Monie TP, Moncrieffe MC, Gay NJ. Structure and regulation of cytoplasmic adapter proteins involved in innate immune signaling. Immunol Rev. 2009 Jan.

[100] Oshiumi H, Matsumoto M, Hatakeyama S, Seya T. Riplet/RNF135, a RING finger protein, ubiquitinates RIG-I to promote interferon-beta induction during the early phase of viral infection. J Biol Chem. 2009 .

[101] Ozato K, Shin DM, Chang TH, Morse HC 3rd. TRIM family proteins and their emerging roles in innate immunity. Nat Rev Immunol. 2008 Nov.

[102] Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P, Salemme FR. High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen. 2001 Dec.

[103] Palese P, Shaw ML. Orthomzxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, editors. Fields virologzy. Philadelphia:Lippincott Williams and Wilkins; 2007.

[104] Parisien JP, Bamming D, Komuro A, Ramachandran A, Rodriguez JJ, Barber G, Wojahn RD, Horvath CM.A shared interface mediates paramyxovirus interference with antiviral RNA helicases MDA5 and LGP2.J Virol. 2009 Jul.

[105] Pichlmair A, Schulz O, Tan CP, Naslund TI, Liljestrom P, Weber F, Reis e Sousa C. RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. Science. 2006 Nov 10.

[106] Pickart CM, Fushman D. Polyubiquitin chains: polymeric protein signals. Curr Opin Chem Biol. 2004 Dec.

[107] Pippig DA, Hellmuth JC, Cui S, Kirchhofer A, Lammens K, Lammens A, Schmidt A, Rothenfusser S, Hopfner KP. The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA. Nucleic Acids Res. 2009 Apr.

[108] Plumet S, Herschke F, Bourhis JM, Valentin H, Longhi S, Gerlier D. Cytosolic 5'-triphosphate ended viral leader transcript of measles virus as activator of the RIG I-mediated interferon response. PLoS ONE. 2007 Mar 14.

[109] Poole E, He B, Lamb RA, Randall RE, Goodbourn S. The V proteins of simian virus 5 and other paramyxoviruses inhibit induction of interferon-beta. Virology. 2002 Nov 10.

[110] Potter JA, Randall RE, Taylor GL. Crystal structure of human IPS-1/MAVS/VISA/Cardif caspase activation recruitment domain. BMC Struct Biol. 2008 Feb 28.

[111] Pymol. DeLano Scientific LLC, 400 Oyster Point Blvd., Suite 213 South San Francisco, CA 94080-1918.

[112] Qin H, Srinivasula SM, Wu G, Fernandes-Alnemri T, Alnemri ES, Shi Y. Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. Nature. 1999 Jun 10.

[113] Ranjith-Kumar CT, Murali A, Dong W, Srisathiyanarayanan D, Vaughan R, Ortiz-Alacantara J, Bhardwaj K, Li X, Li P, Kao CC. Agonist and antagonist recognition by RIG-I, a cytoplasmic innate immunity receptor. J Biol Chem. 2009 Jan 9.

[114] Read RJ. Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr D Biol Crystallogr. 2001 Oct. Erratum in: Acta Crystallogr D Biol Crystallogr. 2003 Feb.

[115] Reymond A, Meroni G, Fantozzi A, Merla G, Cairo S, Luzi L, Riganelli D, Zanaria E, Messali S, Cainarca S, Guffanti A, Minucci S, Pelicci PG, Ballabio A. The tripartite motif family identifies cell compartments. EMBO J. 2001 May 1.

[116] Rhodes DA, de Bono B, Trowsdale J. Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence? Immunology. 2005 Dec.

[117] Rossmann MG. The molecular replacement method. Acta Crystallogr A. 1990 Feb 1.

[118] Rehwinkel J, Tan CP, Goubau D, Schulz O, Pichlmair A, Bier K, Robb N, Vreede F, Barclay W, Fodor E, Reis e Sousa C. RIG-I detects viral genomic RNA during negative-strand RNA virus infection. Cell. 2010 Feb 5.

[119] Salvesen GS, Riedl SJ. Structure of the Fas/FADD complex: a conditional death domain complex mediating signaling by receptor clustering. Cell Cycle. 2009 Sep 1.

[120] Saito T, Hirai R, Loo YM, Owen D, Johnson CL, Sinha SC, Akira S, Fujita T, Gale M Jr. Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. Proc Natl Acad Sci U S A. 2007 Jan 9.

[121] Saito T, Owen DM, Jiang F, Marcotrigiano J, Gale M Jr. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. Nature. 2008 Jul 24.

[122] Satoh T, Kato H, Kumagai Y, Yoneyama M, Sato S, Matsushita K, Tsujimura T, Fujita T, Akira S, Takeuchi O. LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. Proc Natl Acad Sci U S A. 2010 Jan 26.

[123] Schlee M, Hartmann G. The chase for the RIG-I ligand–recent advances. Mol Ther. 2010 Jul.

[124] Schlee M, Roth A, Hornung V, Hagmann CA, Wimmenauer V, Barchet W, Coch C, Janke M, Mihailovic A, Wardle G, Juranek S, Kato H, Kawai T, Poeck H, Fitzgerald KA, Takeuchi O, Akira S, Tuschl T, Latz E, Ludwig J, Hartmann G. Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus.Immunity. 2009 Jul 17.

[125] Schmidt A, Schwerd T, Hamm W, Hellmuth JC, Cui S, Wenzel M, Hoffmann FS, Michallet MC, Besch R, Hopfner KP, Endres S, Rothenfusser S. 5'-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I. Proc Natl Acad Sci U S A. 2009 Jul.

[126] Schroder K, Tschopp J. The inflammasomes. Cell. 2010 Mar 19.

[127] Scott FL, Stec B, Pop C, Dobaczewska MK, Lee JJ, Monosov E, Robinson H, Salvesen GS, Schwarzenbacher R, Riedl SJ. The Fas-FADD death domain complex structure unravels signalling by receptor clustering. Nature. 2009 Feb 19.

[128] Sebastian S, Grütter C, Strambio de Castillia C, Pertel T, Olivari S, Grütter MG, Luban J. An invariant surface patch on the TRIM5alpha PRYSPRY domain is required for retroviral restriction but dispensable for capsid binding. J Virol. 2009 Apr.

[129] Seth RB, Sun L, Ea CK, Chen ZJ. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. Cell. 2005 Sep 9.

[130] Shub DA, Peebles CL, Hampel A. Ribozymes. In: Higgins SJ, HamesbD (editors) RNA Processing: A Practical Approach. IRL Press, Oxford, Vol2, pp211-239.

[131] Song B, Gold B, O'Huigin C, Javanbakht H, Li X, Stremlau M, Winkler C, Dean M, Sodroski J. The B30.2(SPRY) domain of the retroviral restriction factor TRIM5alpha exhibits lineage-specific length and sequence variation in primates. J Virol. 2005 May.

[132] Srimathi T, Robbins SL, Dubas RL, Hasegawa M, Inohara N, Park YC. Monomer/dimer transition of the caspase-recruitment domain of human Nod1. Biochemistry. 2008 Feb 5.

[133] Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr. 2004 Mar. Activation of the beta interferon promoter by unnatural Sendai virus infection requires RIG-I and is inhibited by viral C proteins.

[134] Strähle L, Marq JB, Brini A, Hausmann S, Kolakofsky D, Garcin D. Activation of the beta interferon promoter by unnatural Sendai virus infection requires RIG-I and is inhibited by viral C proteins. J Virol. 2007 Nov.

[135] Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005 May.

[136] Sumpter R Jr, Loo YM, Foy E, Li K, Yoneyama M, Fujita T, Lemon SM, Gale M Jr. Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. J Virol. 2005 Mar.

[137] Takaoka A, Wang Z, Choi MK, Yanai H, Negishi H, Ban T, Lu Y, Miyagishi M, Kodama T, Honda K, Ohba Y, Taniguchi T. DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. Nature. 2007 Jul 26.

[138] Takahasi K, Kumeta H, Tsuduki N, Narita R, Shigemoto T, Hirai R, Yoneyama M, Horiuchi M, Ogura K, Fujita T, Inagaki F. Solution structures of cytosolic RNA sensor MDA5 and LGP2 C-terminal domains: identification of the RNA recognition loop in RIG-I-like receptors. J Biol Chem. 2009 Jun 26.

[139] Takahasi K, Yoneyama M, Nishihori T, Hirai R, Kumeta H, Narita R, Gale M Jr, Inagaki F, Fujita T. Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. Mol Cell. 2008 Feb 29.

[140] Takeuchi O, Akira S. Pattern recognition receptors and inflammation. Cell. 2010 Mar 19.

[141] Takeuchi O, Akira S. Innate immunity to virus infection. Immunol Rev. 2009 Jan.

[142] Triana-Alonso FJ, Dabrowski M, Wadzack J, Nierhaus KH. Self-coded 3'-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase. J Biol Chem. 1995 Mar 17.

[143] Vitour D, Meurs EF. Regulation of interferon production by RIG-I and LGP2 a lesson in self-control. Sci STKE. 2007 May 1.

[144] Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999.

[145] Venkataraman T, Valdes M, Elsby R, Kakuta S, Caceres G, Saijo S, Iwakura Y, Barber GN. Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. J Immunol. 2007 May 15.

[146] Wang Y, Ludwig J, Schuberth C, Goldeck M, Schlee M, Li H, Juranek S, Sheng G, Micura R, Tuschl T, Hartmann G, Patel DJ. Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. Nat Struct Mol Biol. 2010 Jun 27.

[147] Weichenrieder O, Wild K, Strub K, Cusack S. Structure and assembly of the Alu domain of the mammalian signal recognition particle. Nature. 2000 Nov 9.

[148] Weinert C, Grütter C, Roschitzki-Voser H, Mittl PR, Gr´utter MG. The crystal structure of human pyrin b30.2 domain: implications for mutations associated with familial Mediterranean fever. J Mol Biol. 2009 Nov 27.

[149] Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999.

[150] Woo JS, Imm JH, Min CK, Kim KJ, Cha SS, Oh BH. Structural and functional insights into the B30.2/SPRY domain. EMBO J. 2006 Mar 22.

[151] Woo JS, Suh HY, Park SY, Oh BH. Structural basis for protein recognition by B30.2/SPRY domains. Mol Cell. 2006 Dec 28.

[152] Xu LG, Wang YY, Han KJ, Li LY, Zhai Z, Shu HB. VISA is an adapter protein required for virus-triggered IFN-beta signaling. Mol Cell. 2005 Sep 16.

[153] Yan N, Chai J, Lee ES, Gu L, Liu Q, He J, Wu JW, Kokel D, Li H, Hao Q, Xue D, Shi Y. Structure of the CED-4-CED-9 complex provides insights into programmed cell death in Caenorhabditis elegans.

Nature. 2005 Oct 6.

[154] Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, Taira K, Akira S, Fujita T. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat Immunol. 2004 Jul.

[155] Yoneyama M, Fujita T. RNA recognition and signal transduction by RIG-I-like receptors. Immunol Rev. 2009 Jan.

[156] Yuan S, Yu X, Topf M, Ludtke SJ, Wang X, Akey CW. Structure of an apoptosome-procaspase-9 CARD complex. Structure. 2010 May 12.

[157] Zeng W, Sun L, Jiang X, Chen X, Hou F, Adhikari A, Xu M, Chen ZJ. Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. Cell. 2010 Apr 16.

[158] Zhao C, Denison C, Huibregtse JM, Gygi S, Krug RM. Human ISG15 conjugation targets both IFN-induced and constitutively expressed proteins functioning in diverse cellular pathways. Proc Natl Acad Sci U S A. 2005 Jul 19.

[159] Zimm BH. The scattering of light and the radial distribution function of high polymer solutions. J. Chem. Phys. 16, 1948.

[160] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003.