

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

Structural and Functional Analysis of RIG-I Like Helicases

Modulating Spectral Properties of the Green Fluorescent Protein with Nanobodies



Axel Kirchhofer

aus

Mönchengladbach

München, 2009

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfsmittel erarbeitet.

München, am 30.09.2009



.....

(Axel Kirchhofer)

Dissertation eingereicht am:

12.10.2009

1. Gutachter:

Herr Prof. Dr. Karl-Peter Hopfner

2. Gutachter:

Herr Prof. Dr. Roland Beckmann

Mündliche Prüfung am:

25.11.2009

This thesis has been prepared from June 2006 to September 2009 in the laboratory of Professor Dr. Karl-Peter Hopfner at the Gene Center of the Ludwig-Maximilians-University of Munich (LMU).

Parts of this thesis have been published in the following scientific journals:

Kirchhofer A, Helma J, Schmidthals K, Frauer C, Cui S, Karcher A, Pellis M, Muyldermans S, Delucchi CC, Cardoso MC, Leonhardt H, Hopfner KP and Rothbauer U (2009). "Modulating conformation and spectral properties of fluorescent proteins with nanobodies in living cells." Manuscript accepted at Nat Struct Mol Biol.

Pippig DA, Hellmuth JC, Cui S, Kirchhofer A, Lammens K, Lammens A, Schmidt A, Rothenfusser S, Hopfner KP (2009). "The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA." Nucleic Acids Res. **37**(6):2014-25

Myong S, Cui S, Cornish PV, Kirchhofer A, Gack MU, Jung JU, Hopfner KP, Ha T (2009). "Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA." Science. **323**(5917):1070-4.

Gack MU, Kirchhofer A, Shin YC, Inn KS, Liang C, Cui S, Myong S, Ha T, Hopfner KP, Jung JU (2008). "Roles of RIG-I N-terminal tandem CARD and splice variant in TRIM25-mediated antiviral signal transduction." Proc Natl Acad Sci U S A. **105**(43):16743-8

Cui S, Eisenächer K, Kirchhofer A, Brzózka K, Lammens A, Lammens K, Fujita T, Conzelmann KK, Krug A, Hopfner KP (2008). "The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I". Mol Cell. **(2)**:169-79

Parts of this thesis have been presented at an international conference:

Poster and talk "Structural and Functional Analysis of RIG-I Like Helicases" in the plenary session at the Keystone Symposium Pattern Recognition Molecules and Immune Sensors of Pathogens, March 29th – April 2nd 2009 in Banff, Alberta, Canada.

TABLE OF CONTENTS

Part I - Structural and Functional Analysis of RIG-I Like Helicases

1	INTRODUCTION	1
1.1	Pattern Recognition Receptors in Innate Immunity	1
1.2	RIG-I and MDA5: Virus Sensing and RNA Structures Recognized	4
1.3	LGP2: A Regulator of RIG-I and MDA5 Signaling	5
1.4	Downstream of RIG-I and MDA5: Transduction and Regulation of Antiviral Signaling	6
1.5	Inhibition of RLR Signaling by Viral Proteins	8
1.6	Objectives.....	10
2	MATERIALS AND METHODS	11
2.1	Materials	11
2.1.1	Chemicals.....	11
2.1.2	Bacterial Strains	11
2.1.3	Plasmids	11
2.1.4	Media and Supplements	12
2.1.5	Oligonucleotides	13
2.2	Methods.....	14
2.2.1	Bioinformatic Methods	14
2.2.2	Molecular Biology Methods.....	14
2.2.3	Cell Culture Methods	16
2.2.4	Protein Biochemistry Methods.....	17
2.2.5	Functional Protein Assays	18
2.2.6	Peptide SPOT Protein-Peptide Interaction Assay.....	20
2.2.7	Structure Determination by Small Angle X-Ray Scattering.....	21
3	RESULTS	25
3.1	Structure Guided Mutagenesis of the RIG-I Regulatory Domain and RNA Binding Studies.....	25
3.2	Functional Dissection of the Individual RIG-I Domains	27
3.3	Functional Analysis of MDA5 and Its Interaction with <i>Paramyxovirus</i> V-Protein.....	33
3.4	Structural Studies on the MDA5 Regulatory Domain with Small Angle X-Ray Scattering and Homology Modeling.....	38

4	DISCUSSION	42
4.1	The Positively Charged Patch within RIG-I RD is the Recognition Site for 5'-Triphosphate RNA	42
4.2	RIG-I Integrates Two Pathogen-Associated Molecular Patterns	43
4.3	CARDs Play a Dual Regulatory Role	43
4.4	Proposed Model for RIG-I Activation	44
4.5	Hypothetical Modes of Viral Recognition by MDA5.....	45
4.6	V-Protein Interferes with Helicase Activity of MDA5 and Thereby Inhibits Signaling	46
4.7	MDA5 RD Structurally Resembles RIG-I RD.....	46
5	SUMMARY	48

Part II - Modulating Spectral Properties of the Green Fluorescent Protein with Nanobodies

6	INTRODUCTION	51
6.1	Green Fluorescent Protein: From Initial Discovery to its Revolutionary Impact on Live Cell Imaging	51
6.2	Green Fluorescent Protein: Biophysical and Structural Properties	52
6.3	Nanobodies as a Versatile Tool for Specific Protein Targeting	54
6.4	Structural Properties of Nanobodies	56
6.5	Objectives.....	58
7	MATERIALS AND METHODS	59
7.1	Materials	59
7.1.1	Chemicals.....	59
7.1.2	Bacterial Strains	59
7.1.3	Preparation of Minimal Medium for Selenomethionine Expression.....	59
7.2	Methods.....	60
7.2.1	Bioinformatic Methods	60
7.2.2	Protein Biochemistry Methods.....	60
7.2.3	Fluorescence Spectroscopy.....	61
7.2.4	Structure Determination by X-Ray Crystallography	62

8	RESULTS	67
8.1	GFP-Binding Nanobodies Modulate GFP Fluorescence	67
8.2	Purification and Crystallization of the GFP- <i>Enhancer</i> and GFP- <i>Minimizer</i> Complexes.....	69
8.3	Data Collection	71
8.4	Structure Determination and Refinement of the GFP- <i>Enhancer</i> and GFP- <i>Minimizer</i> Complexes .	73
8.5	Structure of GFP- <i>Enhancer</i> and GFP- <i>Minimizer</i> Complexes	74
8.6	Binding of Nanobodies Induces Structural Rearrangements in the Chromophore Environment ...	77
8.7	The <i>Enhancer</i> and the <i>Minimizer</i> Interfere with GFP Chromophore Environment	80
9	DISCUSSION	82
9.1	<i>Enhancer</i> and <i>Minimizer</i> Modulate GFP Intensity in Living Cells	82
9.2	Modulation of GFP Fluorescence with Nanobodies – Future Perspectives.....	84
10	SUMMARY	87
11	REFERENCES	88
12	ACKNOWLEDGMENTS	97
13	CURRICULUM VITAE	98

Part I

Structural and Functional Analysis of RIG-I Like Helicases

1 Introduction

1.1 Pattern Recognition Receptors in Innate Immunity

Key to the establishment of an immune response is the discrimination between “self” and “non self” components within an organism. To this end, specific pathogen-associated molecular patterns (PAMPs) have to be recognized. Vertebrates have evolved two complementary systems to defend themselves against infection by pathogens, the innate and the adaptive immune response. Activation of the innate immune system is the initial response to invading pathogens and in most cases sufficient to clear the infection. Innate immunity is characterized by its ability to recognize a wide range of pathogens such as viruses, bacteria and fungi through a limited number of germline-encoded receptors called pattern recognition receptors (PRRs). In principle, fast evolving pathogens could escape recognition by PRRs by changing their targeted PAMPs. Therefore, the innate immune system recognizes PAMPs that are highly conserved throughout microbial species and essential for viability, such as sugars, flagellin or the cell wall components peptidoglycan and lipopolysaccharide (LPS). An important viral PAMP is double stranded RNA, which is an intermediate in viral replication and is not found in uninfected cells. In case the innate immune system is overwhelmed, the danger signals produced by the innate immune reaction also trigger the adaptive immune response. In adaptive immunity specific recognition is achieved by B and T effector cells which can express an indefinite number of receptors. These are created by somatic gene rearrangement and hypermutation. A detailed discussion of the adaptive immune system is given by Janeway and colleagues (Janeway et al. 2004).

Over the past decade several types of PRRs have been identified (**Table 1**). Probably the best characterized PRRs are Toll-like receptors (TLRs) which are single-pass transmembrane proteins, localizing either to the plasma membrane or to endosomal compartments. In humans more than ten different TLRs with varying ligand-specificities have been described (Gay and Gangloff 2007; Kawai and Akira 2007; O'Neill and Bowie 2007). These receptors are composed of an ectodomain, which consists of multiple leucine-rich repeats (LRRs) that form a characteristic horseshoe fold, a single transmembrane spanning domain and a Toll/IL-1 receptor homology (TIR) domain which faces the cytosol. Current models assume, that TLRs are activated through a ligand-induced dimerization of the receptors which brings the cytosolic TIR domains in close proximity (Jin and Lee 2008). This allows for the recruitment of adaptor molecules, such as MyD88, TRIF and TIRAP which trigger the downstream signaling process. In addition to TLRs, scavenger receptors (SRs) are another type of PRRs which face the extracellular compartment. They are anchored to the cell membrane and are primarily found on macrophages to mediate phagocytosis

(Areschoug and Gordon 2009). Another important function of SRs is that they act as co-receptors to TLRs, recognizing the same microbial patterns and feeding into the analog signaling cascades.

Recognition of PAMPs in the cytosol is mainly carried out by two recently identified classes of proteins, the NOD-like receptors (NLRs) and RIG-I like receptors (RLRs). Structurally, NLRs are multidomain proteins with a tripartite architecture containing a C-terminal region characterized by a series of LRRs, a central nucleotide domain termed the NACHT domain and an N-terminal effector domain (Martinon et al. 2009). The LRR domain has been implicated in ligand sensing, whereas the NACHT domain oligomerizes in an ATP-dependent manner which is necessary for activation of the protein and downstream signaling (Martinon and Tschopp 2004; Faustin et al. 2007). The N-terminal effector domains, which in most cases are either caspase activation and recruitment domains (CARDs) or pyrin domains (PYD), mediate signal transduction to downstream targets. The tertiary structures of PYDs and CARDs are structurally related and are known as the “death fold”, as they are often found in pathways that lead to the activation of caspases or the activation of the transcription factor NF- κ B (Park et al. 2007). Usually, a death domain of one type will interact with another death domain of similar type (e.g. either CARD/CARD – or PYD/PYD-interactions). Therefore the death fold acts as a “molecular velcro” that anchors adaptor and effector proteins to signaling platforms such as the NOD-signalosome or the inflammasome.

The second important class of intracellular PRRs are the RIG-I like helicases. In contrast to the Toll-like PRRs described above, RIG-I like receptors lack repetitive receptor elements such as LRRs which could serve as recognition platform. Therefore, the finding that RIG-I like helicases were involved in viral sensing in the cytoplasm was unexpected (Yoneyama et al. 2004). The most prominent members of this receptor family are RIG-I (retinoic acid inducible gene I) and MDA5 (melanoma differentiation associated antigen 5), which share a similar domain architecture (**Figure 1**). Both proteins are composed of two N-terminal tandem CARDs, which act as adaptor domains analogous to CARDs in NLRs. LGP2 (Laboratory of Genetics and Physiology 2) is the third member of the RLR family. In contrast to RIG-I and MDA5 it lacks the N-terminal adaptor CARDs (**Figure 1**). All three receptors however harbor a central DExD/H box helicase domain which is characterized by seven conserved motifs (I – also known as Walker A, Ia, II also known as Walker B, III IV, V and VI). This domain is implicated in ATP binding and hydrolysis as well as RNA binding. Importantly, intact ATPase activity is essential for downstream signaling (Yoneyama et al. 2004). Recently, three groups independently described a cysteine-rich C-terminus in RIG-I, which is also conserved in MDA5 and LGP2 (Saito et al. 2007; Cui et al. 2008; Takahashi et al. 2008). This domain was found to inhibit RIG-I activation in the absence of viral stimulation.

Furthermore, experiments partially performed within the scope of this thesis identified it as a crucial recognition domain for viral RNA. As it is an important regulator of RIG-I activity the newly-characterized C-terminal domain was termed regulatory domain (RD).

Table 1: Overview of pattern recognition receptors

	Location & function	Protein motifs	References
Surface-PPRs			
Toll-like receptors	Single-pass transmembrane proteins in the extracellular or endosomal membranes. Recognition of various components from viruses, bacteria, protozoa and fungi.	Extracellular leucine rich repeats (LRR), single transmembrane domain, intracellular TIR domain.	(Gay and Gangloff 2007; Kawai and Akira 2007; O'Neill and Bowie 2007)
Scavenger receptors	Receptors found on phagocytotic cells. Enable phagocytosis of pathogens.	No conserved domain architecture. Functional domains include collagenous helices, coiled coils and cysteine-rich regions as well as C-type lectin domains.	(Areschoug and Gordon 2009)
Intracellular PPRs			
NOD-like receptors	Cytosolic receptors inducing a NF κ B and caspase 1 dependent immune response.	N-terminal effector domain (CARD), Pyrin Domain (PYD) or Baculovirus IAP Repeat (BIR)), central NACHT domain and C-terminal LRR.	(Martinon et al. 2009)
RIG-I like receptors	Cytosolic RNA helicases inducing an NF κ B and IRF1 / 3 dependent immune response.	N-terminal tandem CARD, DECH box helicase motif, regulatory domain (RD).	(Yoneyama and Fujita 2009)

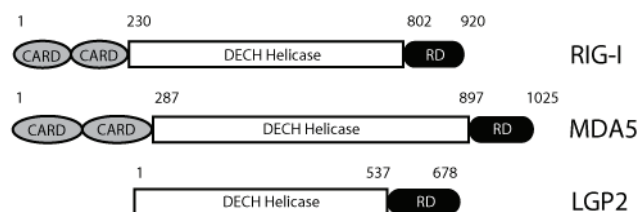


Figure 1: Domain organization of RIG-I like receptors.

RIG-I and MDA5 are composed of two N-terminal tandem CARDS, a central DECH helicase motif and a C-terminal RD domain. LGP2 lacks the N-terminal CARDS but otherwise shows a homologous domain architecture.

The structure of RIG-I RD and the homologous part in LGP2 structure were determined by X-ray crystallography (Cui et al. 2008; Li et al. 2009; Pippig et al. 2009). Both structures are very similar and can be superimposed with only slight differences in the flexible loop regions (**Figure 2**). The RD is a flat domain with a convex and a concave side. It is organized into three antiparallel β -sheets which are connected by two protruding loops. The loops harbor four conserved cysteines which coordinate a zinc atom. The integrity of this zinc binding site is essential for correct folding and functioning of the RD. Analysis of the electrostatic surface potential identified a positively charged patch, which in both RDs localizes to a similar area on the concave side of the molecule (**Figure 6, Figure 19**). Since RIG-I recognizes 5'-triphosphate RNA, which carries a high negative charge, the positive charged patch presumably functions as an RNA recognition site (Cui et al. 2008).

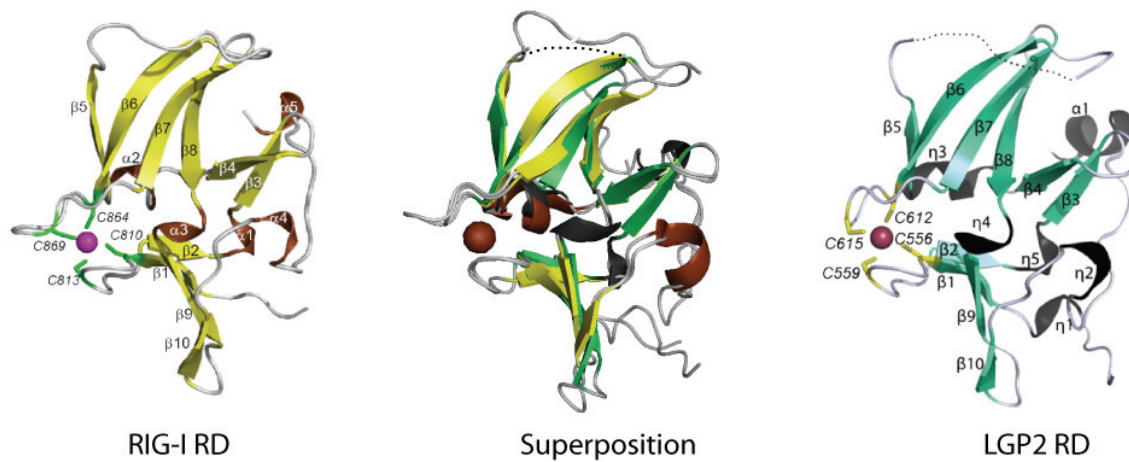


Figure 2: Comparison of RD X-ray structures from RIG-I and LGP2.

Ribbon model of RIG-I and LGP2 RD alone and superimposed with annotated secondary structure, zinc ion (sphere) and zinc-coordinating cysteines (sticks). Adapted from Cui et al 2008 and Pippig et al 2009.

1.2 RIG-I and MDA5: Virus Sensing and RNA Structures Recognized

In order to reproduce themselves once they have entered a cell, viruses must transcribe and replicate their genome. During these processes viral RNA is present in the cytoplasm and could function as a major danger signal for the cell. However in order to recognize this signal, the cell has to be able to discriminate between the viral RNA and the vast amount of self RNA, which in the cytoplasm is mainly constituted of mRNA and tRNA. One epitope specific to certain viral RNAs is the 5'-triphosphate modification, while in contrast, termini from cellular tRNA and mRNA are usually modified or have masking proteins bound (Bowie and Fitzgerald 2007). Two groundbreaking studies have demonstrated that the 5'-triphosphate motif of single stranded RNA (ssRNA) is crucial for RIG-I induced antiviral signaling (Hornung

et al. 2006; Pichlmair et al. 2006). RIG-I recognizes a series of ssRNA negative-strand viruses such as Hepatitis C Virus (HCV) and Measles Virus, which are known to produce 5'-triphosphate RNA intermediates. In addition to 5'-triphosphate RNA as potent stimulator of RIG-I, there is accumulating evidence that double-stranded RNA (dsRNA), another known viral pattern, is also important for RIG-I activation. In particular, RIG-I can be activated with dsRNA *in vitro* as well as *in vivo* in a 5'-triphosphate independent manner (Yoneyama et al. 2004; Cui et al. 2008).

Whereas significant progress has been made over the past years to identify the RIG-I ligands, the physiological ligand for MDA5 remains enigmatic. It has been shown that MDA5 is essential for the production of interferon in response to Picornaviruses, such as *Encephalomyocarditis Virus* (EMCV) (Kato et al. 2006). However, exactly which molecular patterns are recognized remains to be clarified. It is known that 5'-triphosphates on RNA are not necessary for MDA5 specific detection since Picornaviruses harbor a priming protein which masks RNA 5'-ends (Ambros and Baltimore 1978). Moreover, poly I:C, an synthetic dsRNA mimic, is a very potent activator of MDA5 *in vivo* and *in vitro* (Kato et al. 2006). This enzymatically synthesized polymer of undefined length has been known for decades to be a very potent stimulator of interferon (Colby and Chamberlin 1969). The length of the I:C polymer seems to be an important criterion for MDA5 activation. Whereas short poly I:C < 1kbp is primarily sensed by RIG-I, dsRNA > 2kbp is exclusively recognized by MDA5 (Kato et al. 2008). So far, no convincing correlation between molecular features of the synthetic poly I:C and the physiological, MDA5-stimulating PAMPs from *Picornaviridae* could be found.

1.3 LGP2: A Regulator of RIG-I and MDA5 Signaling

LGP2 preferentially binds to dsRNA independent of triphosphate modifications (Murali et al. 2008; Li et al. 2009; Pippig et al. 2009). Binding to dsRNA is mediated through a positively charged patch on the RD. The helicase domain also harbors a dsRNA binding site. It is currently unclear how the RNA binding specificity found *in vitro* correlates with the detection of viral RNA intermediates in infected cells *in vivo*. Due to its lack of CARDs, LGP2 is unable to induce antiviral signaling directly. However, several studies suggest how LGP2 could relay its signal by indirect means. On the one hand, LGP2 seems to have a positive regulatory impact on MDA5 signaling. On the other hand, it functions as a negative regulator of RIG-I mediated signal transduction (Rothenfusser et al. 2005; Komuro and Horvath 2006; Venkataraman et al. 2007). Whereas the negative regulation of RIG-I might be explained by a simple sequestration of dsRNA ligand, the positive regulation of MDA5 signaling suggests that the two proteins interact directly, or through a signaling pathway. However, a direct

interaction between MDA5 and LGP2 has not yet been detected *in vitro*. It is clear that the controversial roles of LGP2 in antiviral signaling are important targets for future studies.

1.4 Downstream of RIG-I and MDA5: Transduction and Regulation of Antiviral Signaling

CARDs act as molecular adaptors to mediate interactions with other CARD containing proteins. This knowledge narrowed down considerably the number of potential downstream adaptors for RLRs. Interferon- β Promoter Stimulator 1 (IPS-1), also known as Mitochondrial Antiviral Signaling (MAVS), Virus-Induced Signaling Adaptor (VISA) and CARD adaptor Inducing IFN- β (CARDIF) were identified by four groups independently to be the adaptors for RIG-I and MDA5 mediated signaling (Kawai et al. 2005; Meylan et al. 2005; Seth et al. 2005; Xu et al. 2005). IPS-1 is located on the outer mitochondrial membrane and consists of an N-terminal CARD followed by a proline rich region (PRR) close to the N-terminus and a single C-terminal transmembrane domain. Activated RIG-I and MDA5 are thought to interact with IPS-1 via their CARDs thereby inducing recruitment of downstream signaling molecules. Several tumor necrosis factors (TNFs) and receptor-associated factors (TRAFs) directly interact with IPS-1 and transmit signals to downstream protein kinases of the inhibitor of NF- κ B kinase (IKK) family. The IKKs are then essential for the activation of the transcription factors IRF-3/7 and NF- κ B. These transcription factors ultimately induce the production of type I interferons and inflammatory cytokines (**Figure 3**).

Activation of RLR signaling has detrimental consequences for the cell and the entire organism. Therefore, it is very important that false positive signaling is prevented. Despite this, very small amounts of viral RNA have to be detected over the vast amount of self RNA. In addition, once the infection has been cleared, the antiviral response needs to be down-regulated promptly. It is therefore obvious that this highly sensitive viral RNA detection system must include a vast array of both positive and negative regulatory proteins. Recent work revealed that post-translational modification, in particular ubiquitination, of RIG-I is important for its signaling capacity. Tripartite motif protein 25 (TRIM25), an E3 ubiquitin ligase, specifically interacts with the first CARD of RIG-I and conjugates Lys-63-linked ubiquitin chains to the Lys172 residue within the second CARD (Gack et al. 2007). These authors demonstrated that TRIM25 activity is needed for direct interaction of RIG-I with IPS-1 and successful signal transduction. Interestingly, MDA5 signaling is independent of TRIM25-mediated ubiquitination. In contrast to TRIM25, which is a positive regulator of RIG-I signaling, the ubiquitin ligase Ring Finger 125 (RNF125) downregulates RIG-I signaling (Arimoto et al. 2007). In this case, Lys-48-linked polyubiquitination in the N-terminal portion of RIG-I and MDA5 induces their proteasomal degradation. Recently, conjugation of ubiquitin-

like ISG15 proteins to RIG-I was described as yet another negative regulatory mechanism (Arimoto et al. 2008).

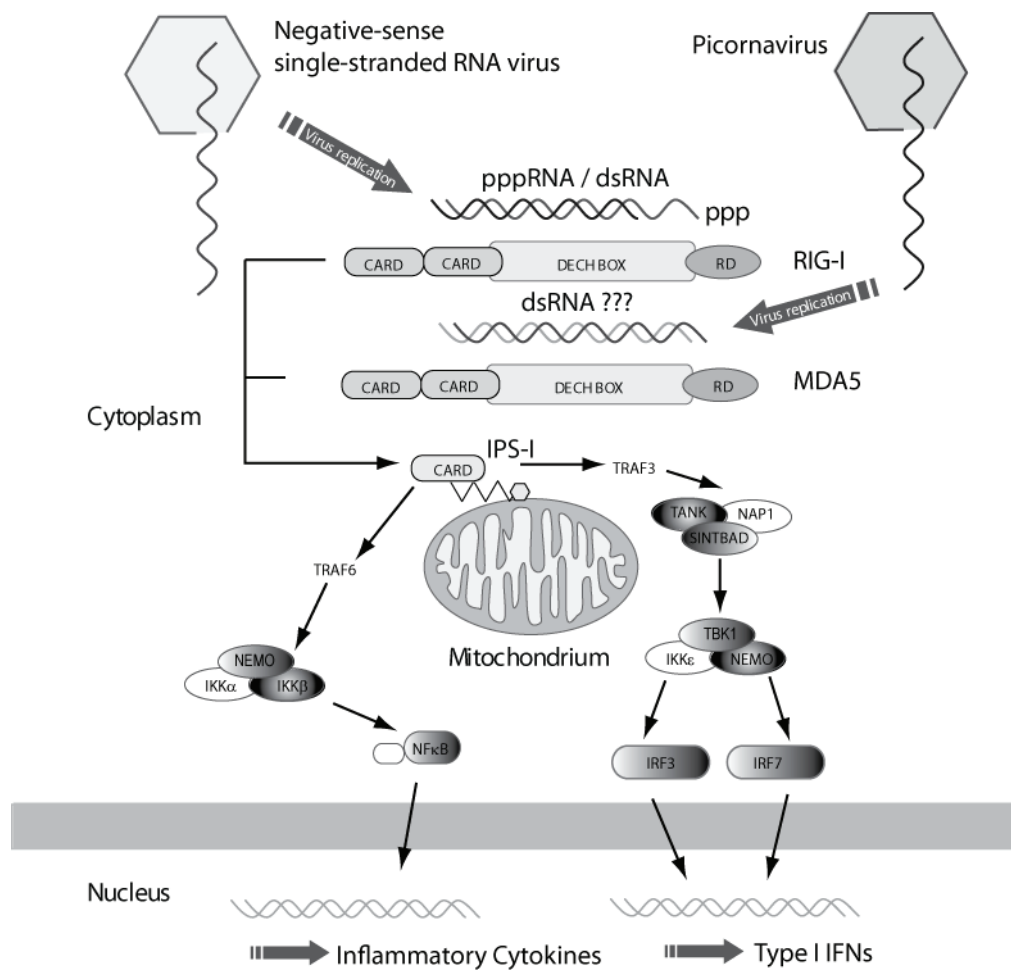


Figure 3: Virus sensing by RLRs in the cytoplasm and subsequent signal transduction.

RIG-I and MDA5 sense RNA from different types of viruses. They interact with the mitochondrial adaptor protein IPS-1 to set off a signaling cascade finally resulting in the production of NFκB and IRF3/7. These transcription factors induce production of inflammatory cytokines and type I interferons (IFNs). Adapted from Yoneyama and Fujita 2009.

Since the CARDS are essential for RLR signaling, these domains are a primary target for the regulation of antiviral signaling. LGP2, which lacks CARDS and thus cannot induce an interferon response *per se*, has been implicated to regulate RIG-I and MDA5 activity (chapter 1.3). A detailed scheme of regulatory factors which influence RLR signaling is shown in **Figure 4**.

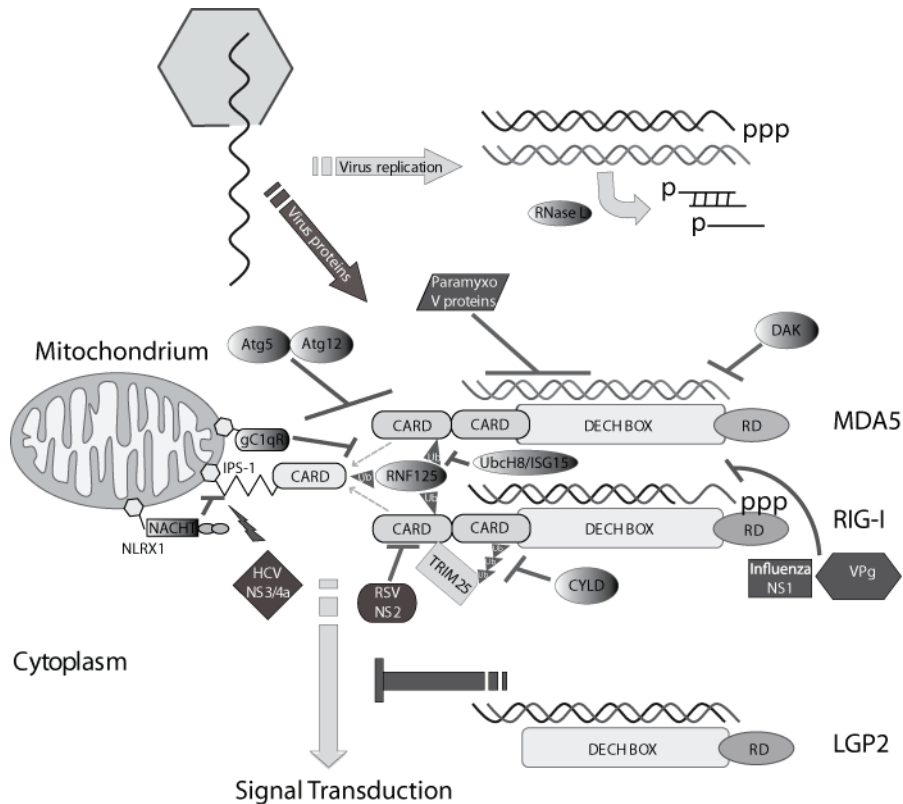


Figure 4: Regulators and viral inhibitors of RLR signaling.

Posttranslational modifications of RIG-I and MDA5 such as ubiquitination by RNF125 or TRIM25 are critical for a tight regulation of the antiviral response. Kinases and other protein factors (fading grey spheres) influence RLR signaling on almost every level of signal transduction. The third helicase LGP2 has both positive and negative regulatory roles on MDA5 and RIG-I, respectively. The RIG-I splice variant (RIG-I SV) is a negative regulator of RIG-I signaling. Several viral proteins (dark grey boxes) interfere with critical components of the RLR signaling pathway.

1.5 Inhibition of RLR Signaling by Viral Proteins

Viruses have evolved a multitude of strategies to evade the host immune response. One mechanism of viral evasion is the interference of viral proteins with components of the host's interferon response. For example, V-proteins of *Paramyxoviridae* have been shown to directly bind and inhibit MDA5-dependent interferon signaling (Andrejeva et al. 2004; Childs et al. 2007). Interestingly, the interaction with V-protein seems to be highly specific for MDA5, with RIG-I signaling remaining unaffected. Moreover, no direct interaction between RIG-I and V-protein could be described. V protein targets MDA5 via its highly conserved cysteine-rich C-terminal domain. Recently it was reported that V-protein binds to a region within the MDA5 helicase domain and inhibits activation of MDA5 by blocking dsRNA binding (Childs et al. 2008). Besides MDA5, V-protein interacts with several other cellular components. For example, V-protein-mediated connection of host proteins, such as STAT1 and STAT2, to the

DDB1-Cul4A-Roc1 E3 ubiquitin ligase complex leads to their specific ubiquitination and proteolytic destruction and thus interferes with the antiviral immune response (Horvath 2004).

In addition to the inhibition of MDA5 by V-protein, several other viral proteins have been shown to specifically counteract central players in RLR signaling. Nonstructural protein 1 (NS1) of influenza A virus interacts with RIG-I and inhibits its function (Guo et al. 2007). NS1, besides its direct influence on RIG-I, also has RNA binding activity and thereby conceals RLR stimulators (**Figure 4**). A similar inhibitory function has been attributed to nonstructural protein 2 (NS2) which interacts with the N-terminal portion of RIG-I (Kaukinen et al. 2006). The viral protease NS3/4 specifically cleaves IPS-1 at Cys508 and inactivates this important signaling adaptor (**Figure 4**). An even more direct strategy to evade recognition by the RLR detection system is the sequestering of viral RNA by viral proteins. Along these lines *Picornaviridae* encode a VPg protein which is covalently linked to triphosphate termini and thereby masks potential RIG-I activators (**Figure 4**) (Ambros and Baltimore 1978).

1.6 Objectives

RIG-I like receptors are PRRs which sense viral RNA in the cytoplasm. Our laboratory was able to solve the crystal structure of the C-terminal RIG-I RD which was shown to be critical for the binding of 5'-triphosphate (Cui et al. 2008). It was, however, unclear how RIG-I recognizes viral RNA at the molecular level. In addition, the function of RD in the context of the full-length protein, in particular how RD synergistically functions together with other parts of the protein, was unknown and needed to be addressed experimentally. The fact that all RLRs harbor a DECH-box helicase motif suggested that these receptors may be endowed with helicase- (unwinding of nucleic acids) or translocation- (movement on nucleic acids) activity. Furthermore, it was in our greatest interest to structurally and functionally compare RIG-I to MDA5, the less-well characterized homologous receptor. Interestingly, MDA5 but not RIG-I function can be inhibited by V-proteins from *Paramyxoviridae*. The molecular mechanism behind this specific inhibition was, however, also completely enigmatic.

Therefore, the aim of this PhD thesis was to functionally dissect individual domains of RIG-I by performing biochemical studies. The binding site of 5-triphosphate ssRNA on RIG-I RD was to be mapped by a structure-guided mutagenesis approach and RNA binding studies. In addition, a detailed analysis of RIG-I ATPase activity was performed to give insights into the functional contribution of individual domains. To complement these biochemical studies, the potential helicase activity of RIG-I was tested on single-molecule level. Furthermore, functional studies of MDA5 and *Paramyxovirus* V-protein aimed to shed light on the molecular basis of the physiological MDA5 ligand and how exactly protein factors are able to inhibit MDA5. Finally, structural analysis of the MDA5 RD should allow a direct comparison to the homologous domain in RIG-I and might help explain the different ligand specificities.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

All common chemicals were obtained from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany) and Sigma (Deisenhofen, Germany), unless otherwise stated. Enzymes and nucleotides for molecular biology were supplied by MBI Fermentas (St. Leon-Rot, Germany). Chromatographic media and columns as well as radiolabel led nucleotides (γ - 32 P-ATP) were purchased from GE Healthcare (Freiburg, Germany). DNA oligonucleotides for cloning were ordered from Eurofins MWG (Munich, Germany). RNA for crystallization and activity assays was acquired from Biomers (Ulm, Germany). Poly I:C for MDA5 activity tests was purchased from Sigma (Deisenhofen, Germany). RIG-I and MDA5 cDNA for molecular cloning of expression constructs were received from RZPD (Heidelberg, Germany).

2.1.2 Bacterial Strains

Table 2: Bacterial strains

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

2.1.3 Plasmids

Table 3: Utilized plasmids

Plasmid	Expression System	Source
pET21b(+)	<i>E. coli</i>	Novagen, Madison USA
pET28b(+)	<i>E. coli</i>	Novagen, Madison USA
pCool pGEX)	(modified <i>E. coli</i>	Ning Zheng (Li et al. 2006)
pFBDM	Insect cells	Imre Berger (Berger et al. 2004)

2.1.4 Media and Supplements

Luria Broth (LB) liquid media as well as LB Agar plates were prepared according to standard protocols (Sambrook et al. 1989). The media was supplemented with the respective antibiotics using stock solutions in 1:1000 dilutions (**Table 4**).

Table 4: Antibiotic stock solutions

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

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

2.1.5 Oligonucleotides

Table 5: DNA oligonucleotides used for molecular cloning

Name	Sequence in 5' to 3' orientation
RIG-I K858A for	GCAGTTTTCAAGTTTTGAAGCAAGAGCAAAGATATTCTGTG
RIG-I K858A rev	CACAGAATATCTTTGCTCTTGCTTCAAACCTTGAAAACCTGC
RIG-I K888A for	CATTTGAGATTCCAGTTATAGCAATTGAAAGTTTTGTGGTGG
RIG-I K888A rev	CCACCACAAAACCTTCAATTGCTATAACTGGAATCTCAAATG
RIG-I H830A for	GTAAGAGTGATAGAGGAATGCGCTTACACTGTGCTTGGAGATG
RIG-I H830A rev	CATCTCCAAGCACAGTGTAAGCGCATTCTCTATCACTCTTAC
RIG-I I875A for	CAGCCATGACTGGGGAGCCCATGTGAAGTACAAG
RIG-I I875A rev	CTTGACTTTCACATGGGCTCCCCAGTCATGGCTG
RIG-I D836A for	CATTACACTGTGCTTGGAGCTGCTTTTAAGGAATGCTTTG
RIG-I D836A rev	CAAAGCATTCTTAAAAGCAGCTCCAAGCACAGTGTAATG
RIG-I K807A for	CCTGATAAGGAAAATAAAGCACTGCTCTGCAGAAAGTGC
RIG-I K807A rev	GCACTTTCTGCAGAGCAGTGCTTTATTTTCCTTATCAGG
MDA5 Sal1 for	AAAAAAGTCGACATGTGCAATGGGTATTCCACAGAC
MDA5 Not1 r 6his	AAAAAAGCGGCCGCTCAATGATGATGATGATGATGATGATCCTCATCACTAAATAAACAGCAT TC
MDA5 Nhe1 287 for	AAAAAAGCTAGCATGGGAAGTGATTCAGATGAAGAGAA
MDA5 BamH1 897 rev	AAAAAAGGATCCTCACTTGTAAATGCTTGGCAATATTTCTCTTG
MDA5 BamH1 1025 rev	AAAAAAGGATCCCTAATCCTCATCACTAAATAAACAGCATTC
MDA5 897 Nde1 for	TGCCGCGCGGCAGCCATATG AAGAATAACCCATCACTAATAACTTTCC
MDA5 Not1 rev	TGGTGCTCGAGTGCGGCCGCTAATCCTCATCACTAAATAAACAGCATTC
V measles Nde1 49 for	AAAAAACATATGACCTGCAGGGAAGAGAAGGCA
V measles NotI rev	AAAAAA GCGGCCGCTTATTCTGGGATCTCGGGGAGG

Table 6: RNA Oligonucleotides used for functional assays

Name	Origin	Sequence in 5' to 3' orientation
pppRVL	<i>in vitro</i> trans- cription	pppCUUAACAACCAGAUCAAAGAAAAAACAGACAGCGUCAUUGGCAGAGCAAAAA UGU
dsRNA for *	synthetic	ACGCUUAACAACCAGAUCAAAGAAAAAACAGACAGCGUCAUUGGCAGAGC
dsRNA rev*	synthetic	GCUCUGCCAUUGACGCGUCUGUCUGUUUUUUCUUUGAUCUGGUUGUUAAGCGU
3'P	synthetic	ACGCUUAACAACCAGAUCAAAGAAAAAp
short I:C for *	synthetic	AAAAA CCCCCCACCCCCC AAAAA
short I:C for mismatch *	synthetic	AAAAA CCCUCCCACCCUCCC AAAAA
short rev	I:C synthetic	UUUUUU IIIIIIIIIIIII UUUU

*Complementary single stranded oligos were annealed by heating them to 95°C for 5 min followed by gradually cooling them down to 20°C with a temperature decrease of 0.5°C per min using a thermocycler.

2.2 Methods

2.2.1 Bioinformatic Methods

2.2.1.1 *Sequence Alignments*

Multiple sequence alignments were built with ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/>) and visualized using Esprit (<http://esprit.ibcp.fr/ESPript/ESPript/index.php>).

2.2.1.2 *Calculation of Protein Parameters*

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

2.2.1.3 *Structure Visualization and Analysis*

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

2.2.1.4 *Structural Homology Modeling*

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

2.2.2 Molecular Biology Methods

2.2.2.1 *Oligonucleotide Design and Polymerase Chain Reaction (PCR)*

PCR Primers for cloning genes of interest were designed using GeneRunner (Hastings software, USA). In general, the oligonucleotides used had a region complementary to the amplified DNA end of 20-25 nucleotides with a melting temperature (T_m) between 68°-72°C, and an attached restriction endonuclease site with a 6 poly(A) overhang to assure efficiency of restriction endonuclease cleavage. RIG-I and MDA5 cDNA was amplified by PCR using Phusion Flash Master Mix (Finnzymes, Espoo, Finland). Primer concentration was 0.5 μ M each, and approximately 1 ng of template DNA was added to each reaction. The final volume of each reaction was 20 μ l. Each thermocycling program used 30 cycles, with times and

temperatures of denaturation, annealing and elongation varied to achieve optimal amplification. PCR products were purified from agarose gels using a Gel Extraction Kit (Metabion, Martinsried, Germany).

2.2.2.2 Site-Directed Mutagenesis

Point mutations were introduced by PCR using the quick change protocol (Kunkel 1985). Two synthetic oligonucleotide primers containing the desired mutation were extended during temperature cycling with Phusion polymerase. Incorporation of the oligonucleotide primers generated a mutated plasmid containing staggered nicks. A typical thermocycling procedure consisted of a 15 sec denaturation step at 98°C, a 30 sec annealing step at 58°C and a 5 min extension step at 72°C. A complete run typically consisted of 20 repetitive cycles. Following temperature cycling, the product was treated with Dpn I to remove methylated template DNA and the reaction mixture was then transformed into competent XL1 blue cells.

2.2.2.3 Restriction Cleavage and Ligation

DNA was digested using restriction endonucleases and corresponding buffers (Fermentas, St. Leon-Rot, Germany) as recommended by the manufacturer. Digested products were purified using a Gel Extraction Kit (Metabion, Matrinsried, Germany). Cleaved vector DNA was additionally treated with calf intestine alkaline phosphatase (Fast AP, Fermentas, St. Leon-Rot, Germany) to prevent self-ligation. For ligation, a five- to tenfold excess of the digested DNA fragment was incubated with linearized vector and T4 DNA-ligase (Fermentas, St. Leon-Rot, Germany) in a 20 µl reaction according to the manufacturer's instructions.

2.2.2.3 Transformation of E. coli and Isolation of Plasmid DNA

Transformation of plasmid DNA into competent *E. coli* cells was performed by mixing 100 µl of competent cells with 10 µl of the ligation reaction or 1 µl of purified plasmid DNA (40 ng/µl) and incubating on ice for 20 min. Cells were heat-shocked at 42°C for 45 sec and immediately afterwards chilled on ice. 1000 µl of fresh LB medium was added, followed by incubation at 37°C for 45 min in a shaking incubator. Cells were plated on LB agar plates containing the respective antibiotics and incubated at 37°C overnight. Plasmid DNA was isolated from a 5 ml overnight culture using the Plasmid Extraction Kit (Metabion, Matrinsried, Germany). DNA-sequencing of wild-type and mutant clones was performed by Eurofins MWG (Munich, Germany).

2.2.2.4 Bacmid Preparation for Expression in Insect Cells

Plasmid DNA of expression constructs was transformed into chemical competent DH10MultiBac cells as described previously (Section 2.2.2.3). After heat shock at 42°C and

cooling on ice, cells were incubated in a shaking incubator for 6-12 hours at 37°C then plated on agar plates, containing the appropriate antibiotics (Kanamycin, Gentamycin, Tetracyclin) plus X- α -galactose (100 μ g/ml) and IPTG (40 μ g/ml). Plates were incubated at 37°C for 48 h. Successful integration of the gene of interest into the bacmid was assessed by blue / white screening. White colonies were used to inoculate an overnight shaking culture of 100 ml LB plus respective antibiotics. Cells were harvested by centrifugation and the bacmid isolated using the Midi-Prep Kit (Qiagen, Hilden, Germany), following the instructions provided by the manufacturer.

2.2.2.5 Electrophoretic Separation of DNA

DNA was separated in horizontally poured 1 % (w/v) Agarose/1x TAE gels containing 0.7 μ g/ml ethidium bromide (Roth, Karlsruhe, Germany) in the appropriate gel chamber filled with 1x TAE buffer (40 mM Tris, 20 mM Acetic Acid, 1 mM EDTA). Before loading, samples were mixed with 6X Loading dye (Fermentas, St. Leon-Rot, Germany). DNA was visualized using a standard ultraviolet transilluminator (λ =254 nm, Eagle Eye, Stratagene, LaJolla, USA).

2.2.3 Cell Culture Methods

2.2.3.1 Protein Expression in *E. coli*

For the over-expression of recombinant proteins, competent *E. coli* Rosetta (DE3) cells (Novagen, Schwalbach, Germany) were transformed with plasmid DNA carrying the gene of interest. Cells were grown at 37°C in LB medium in the presence of the appropriate antibiotics to an OD₆₀₀ of 0.6-0.8. Protein expression was induced by the addition of IPTG (Roth, Karlsruhe, Germany) to a final concentration of 0.5 mM. After further growth overnight at 18°C, cells were harvested by centrifugation at 4°C. Cell pellets were flash frozen in liquid nitrogen and stored at -80°C.

2.2.3.2 Protein Expression in Insect Cells

Freshly diluted High Five insect cells (Invitrogen, Karlsruhe, Germany) were seeded in a 6-well tissue culture plate (0.5 x 10⁶ cells per well). Typically, cells were infected with 5 μ g of bacmid DNA mixed with 3 μ l of FuGENE HD Transfection Reagent (Roche, Penzberg, Germany) according to the manufacturer's protocol. After incubation of the infected cells for 48-60 h at 27.5 °C, the supernatant (viral generation 0 - V₀) was collected. The viral titer was amplified by infecting 50 ml of freshly resuspended High Five insect cells at a concentration of 0.5 x 10⁶ cells / ml with 3ml of V₀. Cells were cultured for 72 - 96 h in 500 ml flasks at 27.5 C with shaking at 85 rpm. During the incubation the cell density was monitored at regular intervals and cells were split below 2 x 10⁶ cells/ml until cell proliferation arrested.

The supernatant (V_1) was harvested by centrifugation and the cell material used for small scale purification. Large scale expressions were performed by incubating 1-5 L of freshly resuspended High Five insect cells (cell density: 0.5×10^6 cells/ml) with 20 - 50 ml of higher generation virus ($V_{\geq 1}$). Cells were cultured in 5L flasks (maximum 1L per flask) at similar conditions as described for viral titer amplification. 48 h after cell proliferation had ceased, cells were harvested by centrifugation, flash frozen in liquid nitrogen and stored at -80°C . The supernatant containing high viral titers was utilized for further expression.

2.2.4 Protein Biochemistry Methods

2.2.4.1 Protein separation by SDS-PAGE

Protein samples were analyzed by SDS-PAGE using 15% (v/v) polyacrylamide gels using the vertical Mini-PROTEAN 3 System (BIO-RAD, Munich, Germany). Before loading, samples were mixed with 4x loading dye and heated at 95°C for 2 min. After electrophoresis in electrophoresis buffer (190 mM Glycine, 50 mM tris, 0.1% (w/v) SDS), gels were stained with Coomassie staining solution (50% (v/v) ethanol, 7% (v/v) acetic acid, 0.2% (w/v) Coomassie Brilliant blue R-250) and destained with water.

2.2.4.2 Protein Purification

All buffers and their components used for protein purification are listed in **Table 7**

For purification, cell pellets were resuspended in lysis buffer and disrupted by sonication. Cell debris was removed by centrifugation. His-tagged constructs were first purified using Ni-NTA Agarose (Qiagen, Hilden, Germany) packed in a gravity flow cartridge (BIO-RAD, Munich, Germany). After loading of soluble extracts, the resin was washed with Ni-NTA washing buffer. Protein was then eluted with Ni-NTA elution buffer. GST-tagged SV5 V-protein complexed with His-tagged MDA5 was further purified in a second affinity purification step using a GST-sepharose column (GE Healthcare, Freiburg, Germany). After capturing and washing with GST washing buffer, the complex was eluted with GST elution buffer. According to the theoretical pI (see chapter 2.2.1.2) all proteins were then further purified using anion- or cation- exchange columns (**Table 8**). The ion exchange column was equilibrated with low salt buffer and after loading, the protein was eluted with a gradient of 20 column volumes from low- to high salt buffer. Peak fractions were pooled, concentrated with centrifugal devices (Amicon® Ultra, Millipore, Billerica, USA) and loaded onto a Superdex S200 26/60 size exclusion chromatography column (GE Healthcare, Freiburg, Germany) previously equilibrated with gel filtration buffer. Peak fractions were concentrated to greater than 5 mg/ml, flash frozen in liquid nitrogen and stored at -80°C .

Table 7: Protein purification buffers

Buffer	Composition
Lysis buffer	50 mM Tris 7.5, 400 mM NaCl, 20 mM Imidazol, 10 mM β -mercaptoethanol, 1 mM PMSF
Ni-NTA washing buffer	50 mM Tris 7.5, 400 mM NaCl, 20 mM Imidazol, 10 mM β -mercaptoethanol,
Ni-NTA elution buffer	50 mM Tris 7.5, 100 mM NaCl, 200 mM Imidazol, 10 mM β -mercaptoethanol,
GST washing buffer	50 mM Tris 8.0, 150 mM NaCl, 1mM DTT,
GST elution buffer	50 mM Tris 8.0, 150 mM NaCl, 20 mM glutathione, 1mM DTT
Low salt buffer	50 mM Tris 7.5, 100 mM NaCl, 1mM DTT, 100 μ M ZnCl ₂
High salt buffer	50 mM Tris 7.5, 1000 mM NaCl, 1mM DTT, 100 μ M ZnCl ₂
Gel filtration buffer	50 mM Tris 7.5, 150 mM NaCl, 5mM DTT, 100 μ M ZnCl ₂

Table 8: Purification strategy for used protein constructs

Protein construct	Purification steps
RIG-I full length	Ni-NTA, Q-sepharose, S200 gelfiltration
RIG-I helicase domain	Ni-NTA, Q-sepharose, S200 gelfiltration
RIG-I- Δ -CARD	Ni-NTA, Q-sepharose, S200 gelfiltration
RIG-I SV	Ni-NTA, Q-sepharose, S200 gelfiltration
MDA5-FL	Ni-NTA, Q-sepharose, S200 gelfiltration
MDA5- Δ -CARD	Ni-NTA, Q-sepharose, S200 gelfiltration
SV5 V-protein MDA5 complex	Ni-NTA, GST-sepharose, Q-sepharose, S200 gelfiltration
MDA5 RD	Ni-NTA, S-Sepharose, S200 gelfiltration
RIG-I RD	Ni-NTA, S-Sepharose, S200 gelfiltration

2.2.5 Functional Protein Assays

2.2.5.1 ATPase Assays

The RNA-stimulated ATPase assays were performed in a 50 μ l reaction volume containing 100 mM Tris pH 7.5, 100 mM NaCl, 5 mM MgCl₂, 100 μ g/ml bovine serum albumin (BSA), 1 mM DTT, 0.1 - 5 mM ATP, 20 nM γ -³²P-ATP (3000 Ci/mmol, Amersham Bioscience) and 0.1 μ M stimulatory RNA (in case of RIG-I: pppRVL, in case of MDA5 poly I:C or other synthetic RNA). After equilibration of the sample to 37°C, the reaction was started by addition of protein (typically 0.1 μ M). A time course of ATP hydrolysis was made by removing 1 μ l aliquots in constant intervals and directly spotting onto a thin layer chromatography polyethyleneimine-cellulose plate (Merck, Darmstadt). Rapid air-drying of the spotted reaction mixture ensured that the reaction was stopped immediately. The thin-layer chromatography (TLC) plate was developed in 0.5 M LiCl / 1.0 M formic acid and subsequently air dried. The amount of educt

and product (ATP and liberated γ - ^{32}P , respectively) were visualized with a Phosphorimager (Amersham Bioscience) and quantified using Image Quant software. Initial reaction velocities were determined by plotting the amounts of hydrolyzed γ -phosphate against time and performing a linear regression with Origin analysis software. Enzymatic efficiencies were calculated by fitting the initial velocities by non-linear regression according to a modified Michaelis Menten model using the following equation:

$$activity = [enzyme] \frac{k_{cat} / K_M \times [ATP]}{1 + [ATP] / K_M}$$

2.2.5.2 RNA Binding Assay by Fluorescence Anisotropy Measurements

Fluorescent anisotropy measurements were performed using a FluoroMax-P fluorimeter (HORIBA Jobin Yvon, Unterhachingen, Germany) equipped with a Glan-Thompson prism polarizer and connected to a Haake F3 thermostat (Thermo Haake, Karlsruhe, Germany). Typically, 1.5 ml of buffer (20 mM Tris pH 7.5, 150 mM NaCl, 10 μM ZnCl_2 , 1mM DTT) and 50 nM fluorescently labeled RNA (*in vitro* transcribed pppRVL with incorporated Alexa Fluor 488-5-UTP) were pre-equilibrated in a quartz cuvette at 12°C. The protein was added in a stepwise manner and briefly mixed by magnetic stirring. After 3 min re-equilibration, anisotropy data were collected using an excitation wavelength of 492 nm and monitoring the emission at 512 nm. The band pass was 5 nm for excitation and 5 nm for emission. A maximum number of ten trials were performed until minimal deviation of the signal was reached. The data was fitted according to a one site binding model by least square fitting using the Origin data analysis software:

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

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

2.2.5.3 Dot Blot Double Filter Binding Assays

MDA5 binding to poly I:C was measured by filter binding assays on a double layer of nitrocellulose and nylon filters (Manzan et al. 2004). Prior to filtration, the nitrocellulose filter was activated by soaking it in 0.4 M KOH for 10 min followed by extensive rinsing with water. Both membranes were subsequently equilibrated in assay buffer (20 mM Tris pH 7.5, 150 mM NaCl, 10 μM ZnCl_2 , 1mM DTT) for 10 min. Poly I:C (Sigma, Deisenhofen, Germany) was 5' ^{32}P labeled by T4 Polynucleotide kinase (MBI Fermentas) according to the manufacturer's instructions. A concentration series ranging from 0 μM to 5 μM protein was incubated with 50

nM poly I:C for 20 min at 4°C and then filtered through a double layer of a nitrocellulose filter (PROTRAN BA85, Schleicher & Schuell) and a nylon filter (Roti-Nylon plus, Roth; nitrocellulose on top of nylon membrane) in a 96 well dot blot apparatus (Schleicher & Schuell). Proteins and protein poly I:C complexes bind to the nitrocellulose filter, while the remaining free oligonucleotides bind to the nylon filter allowing complete recovery of labeled poly I:C. For quantification the radioactivity was recorded using a STORM Phosphorimager system (Amersham). The ratio of protein-bound to unbound RNA was determined using Image Quant software.

2.2.6 Peptide SPOT Protein-Peptide Interaction Assay

Overlapping MDA5 peptides of 12 amino acids in length were synthesized on a nitrocellulose membrane in an array of spots with an offset of 2 amino acids using an automatic MultiPep synthesis robot (Intavis, Cologne, Germany) following the manufacturer's instructions. Two identical peptide arrays covering the full-length MDA5 protein were synthesized in parallel. Potential V-protein MDA5 peptide interactions were assayed by Far Western blotting. All buffers and their components used for Far Western blotting are listed in **Table 9**. After SPOT synthesis, the nitrocellulose membrane was washed three times with TBS-T and then blocked for 2-3 h with TBS-T Milk. Myc-tagged GST-V-protein or alternatively control protein (myc-GST, kindly provided by Natalie Hiller) was added to the blocking solution (final concentration 1 µM) and incubated at 4°C overnight shaking. The membrane was consecutively washed with TBS-T, TBS-T NaCl, TBS-T Triton, and again TBS-T for 5 min shaking. The membrane was then incubated with the primary myc specific antibody (Myc A-14 sc-789, rabbit, Santa Cruz Biotechnology, Santa Cruz, USA) diluted 1:2000 in TBS-T Milk and subsequently washed as described before. After incubation with the secondary antibody (Peroxidase AffiniPure Goat Anti-Rabbit IgG, Jackson ImmunoResearch, Suffolk, UK) in a 1:5000 dilution and similar washing procedures, the membrane was developed using ECL Western Blotting Detection Reagent (GE Healthcare, Freiburg, Germany) following the manufacturer's instructions. Detection of luminescent signal was performed on a Luminescence Image Analyzer LAS-3000 (Fuji Film, Woodbridge, USA).

Table 9: Buffers used for Far Western detection

Buffer	Composition
TBS-T	20 mM Tris pH 7.6, 140 mM NaCl, 0,1 % (v/v) Tween 20
TBS-T NaCl	(20 mM Tris pH 7.6, 500 mM NaCl, 0,1 % (v/v) Tween 20)
TBS-T Triton	TBS-T + 0.5% (v/v) Triton X-100
TBS-T Milk	TBS-T + 5% (w/v) milk powder

2.2.7 Structure Determination by Small Angle X-Ray Scattering

Small angle X-ray scattering allows gathering of structural information of proteins and other biomolecules without the need for high ordered crystals as needed for X-ray crystallography. The method has the advantage of making flexible parts of the protein, which usually hinder crystallization, accessible for structural analysis. Furthermore, SAXS measurements of protein samples are performed in solution, thus closer reflecting the natural conditions in which proteins occur *in vivo*. SAXS data can be used to calculate an *ab initio* protein model. However, the structures obtained are of relatively low resolution (maximum 15Å) and the obtained solutions for shape reconstructions are not always unambiguously interpretable. Comparison with existing high resolution data such as partial crystal structures is therefore advisable to further evaluate the obtained model. Through this combination with X-ray crystallography, SAXS is a powerful tool which may provide more accurate and complete models of protein structures, conformation, interactions and assemblies in solution (Putnam et al. 2007).

In a SAXS experiment the X-ray wave is scattered mainly by the electrons surrounding the atoms of the examined biomolecule (protein). As the contribution of the solvent scattering is relatively high, SAXS experiments on proteins require separate measurements of the scattering from sample and solvent. As SAXS is performed in solution, the proteins are spatially averaged and no diffraction pattern can be observed. Instead the scattering image is recorded as a function of the momentum transfer $s = 4\pi\lambda^{-1} \sin(\theta)$ where 2θ is the scattering angle. At the lowest resolution, SAXS scattering is dictated by a single size parameter, the radius of gyration (R_G). The R_G is the square root of the average squared distances of each scatterer from the center of the particle that is scattering X-rays and therefore a measure for the size of the particle. For well-behaved samples with no interparticle interference or aggregation the scattering can be related by the Guinier approximation,

$$I(s) \cong I_0 e^{-\frac{1}{3}R_G^2 s^2}$$

where I_0 is the intensity at the scattering amplitude $s = 0$. The Guinier plot, where $\ln(I(s))$ is plotted versus s^2 should give a linear function with I_0 as intercept and a slope that can be used to calculate R_G . Only very homogenous samples with low intermolecular attractive forces will give a linear Guinier plot. Therefore, it is also a useful tool to test sample quality.

Whereas analysis of low s -values gives information about the particle dimension, analysis of high s -values yields details regarding the molecular shape. For a folded macromolecule the intensity of the scattering falls off with Porod's law: $I(s) \propto s^{-4}$

The Kratky plot ($s^2I(s)$) as a function of s , which is deduced from the Porod's law and which can be calculated directly from the scattering curve, provides an excellent tool for evaluating the folding status of samples. For folded domains, which have intensities that fall off as s^{-4} , the Kratky plot yields a peak roughly shaped like a parabola. Unfolded peptides lack the characteristic folded peak and are linear with respect to s in the large s region.

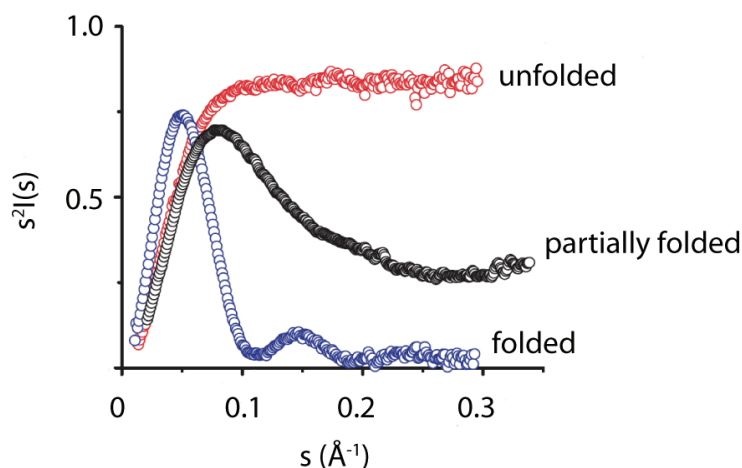


Figure 5: The Kratky plot gives information about protein folding.

Globular macromolecules have bell-shaped curves. Unfolded proteins lack this peak and have a plateau in the larger s range. Adapted from Putman et al 2007.

After the quality of the protein sample has been critically assessed, the the scattering profile $I(s)$ is fourier transformed into the pair distribution function (**Figure 19**) according to the following equation:

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin sr}{sr} dr$$

The pair distribution function can be understood as a two-dimensional Patterson function that illustrates the frequency with which vectors of the length r connect two volume elements within the molecule. The pair distribution function is a real space representation and therefore much more intuitive than the primary scattering curve. The shape of the pair distribution function provides information about the shape of the molecule. Theoretically, the $p(r)$ is zero at $r = 0$ and at $r \geq D_{\max}$, where D_{\max} corresponds to the maximum linear dimension of the scattering particle. D_{\max} cannot be directly calculated from the scattering data but must be estimated for a given molecule. In practice an iterative process is used, with multiple D_{\max} values chosen and the resulting $p(r)$ functions evaluated for their fit to the experimental data.

As described above, several sample parameters can be derived in a straightforward manner directly from the scattering data. In a computationally more demanding step the three-dimensional reconstruction of the particle from the two-dimensional scattering curve is also

feasible. Fitting the data leads to a multidimensional minimization problem that can be however numerically solved. To achieve reasonable models it is helpful to add physical restraints, like imposing a uniform density on the interior of the molecule. The program GASBOR models a protein structure by comparing thousands of configurations of a chain of dummy residues where each dummy correspond to a single amino acid in the protein of interest (Konarev et al. 2006). The structure is approximated by iterative rounds of simulated annealing with additional penalties for non-protein-like density. To improve model quality the results of several individual GASBOR runs should then be aligned and averaged for a final shape. Through comparison of different models and assessment of their capability, an evaluation of the reliability of the final model is possible. In a perfect monodisperse system all runs should converge on a similar structure.

2.2.7.1 Sample Preparation

MDA5-RD samples were purified as described (see chapter 2.2.4.2). The final gel filtration buffer contained 5 mM DTT in order to reduce radiation damage. The purified protein was concentrated using centrifugal devices (Amicon® Ultra, Millipore, Billerica, USA) and aliquots were taken at 2, 5 and 10 mg/ml. A sufficient volume of gel filtration buffer from the same preparation used for the final purification step was stored for reference measurements (**Table 7**).

2.2.7.2 Sample Measurement

MDA5-RD samples were measured at the DESY beamline X33 (Deutsches Elektronen-Synchrotron Hamburg, Germany). All measurements were performed at room temperature after centrifugation in a cooling centrifuge. As a primary quality check, samples were analyzed for protein aggregation by calculating the Guinier plot using the PRIMUS software (Konarev et al. 2006). To evaluate possible radiation damage, MDA5-RD was subjected to three consecutive exposures of different length (6 seconds, 60 seconds and 6 seconds). During the following data acquisition, MDA5-RD samples were measured at increasing protein concentrations. A blank measurement with sample buffer was taken before and after each protein measurement.

2.2.7.3 Data Analysis and *Ab Initio* Modeling

Initial analysis of the primary scattering data was performed using the PRIMUS software package (Konarev et al. 2006). Scattering contributions of buffer were eliminated by subtracting the averaged blank measurements. The protein scattering intensity was scaled according to the respective concentration and subsequently merged in one dataset. The radius of gyration (R_G) was calculated using the linear region Guinier plot. Correct folding of

the protein was evaluated in the Kratky plot. During iterative improvement of the pair distribution function ($p(r)$) using the program GNOM the correct maximum particle diameter (D_{\max}) was estimated by evaluating the resulting R_G value, the $I(s)$ -fit and the shape of the $p(r)$ -distribution. *Ab initio* modeling of the MDA5-RD solution structure was done with GASBORp (Svergun et al., 2001). More than 10 identically calculated models were aligned and averaged using DAMAVER and SUPCOMB (Konarev et al. 2006).

3 Results

3.1 Structure Guided Mutagenesis of the RIG-I Regulatory Domain and RNA Binding Studies

Through a limited proteolysis approach our laboratory identified a stable part of the C-terminal RD of RIG-I. This was amenable to crystallization and its three dimensional structure was determined by X-ray crystallography. Preliminary results from Sheng Cui hinted to a potential role of RD in recognition of the 5'-triphosphate motif of viral RNA. Analysis of the electrostatic surface potential indicated a positively charged patch on the concave side of RD which might be important for RNA recognition. Since RIG-I RD resisted crystallization with RNA ligands, structure-guided mutagenesis was performed on several basic residues within the positively charged cleft to reveal the influence of the individual mutations on RNA binding. The mutants were subsequently used for RNA binding studies and their binding capabilities were compared to that of RIG-I RD wild type protein.

The electrostatic potential of RIG-RD was calculated using the APBS plugin for PyMol (**Figure 6**). Primarily, basic residues in the center of the positively charged pocket were targeted for site specific mutagenesis. In addition, two control residues located on the convex side of RIG-I RD, and therefore should not affect RNA binding, were also targeted. The respective residues (K858, K888, I875, H830, D836, K807 and I875) were all mutated to alanine.

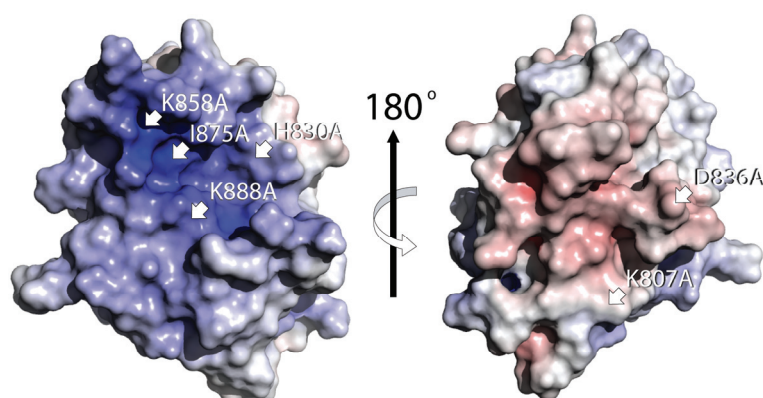


Figure 6: Analysis of the RIG-I RD crystal structure and introduced mutations.

Electrostatic surface potential (ranging from blue = 9 kT/e to red = -9 kT/e), displayed in two different views. Sites of mutated residues are annotated.

Subsequently, the six mutants were expressed and purified in parallel (**Figure 7**) using the previously established protein production protocols for the RIG-I RD wildtype (Cui et al. 2008). Expression of all mutants resulted in yields comparable to the wild type construct.

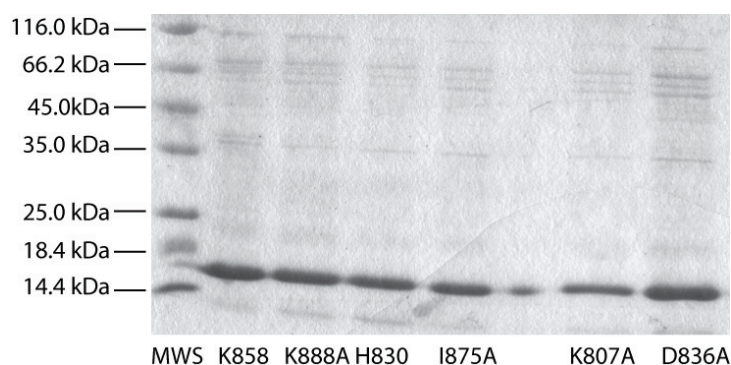


Figure 7: Expression and purification of RIG-I RD mutants.

SDS-PAGE analysis and Coomassie staining of RIG-I RD mutants after purification. Molecular weight standard (MWS) with respective sizes and respective mutations are annotated.

The RNA binding affinities of the indicated RIG-I RD mutants were determined and compared to RD WT and RIG-I full length protein using a fluorescence anisotropy assay. For this, the RNA was fluorescently labeled by incorporating Alexa Fluor 488-UTP into the RVL sequence during the *in vitro* transcription reaction. Titration of increasing amounts of protein into a constant concentration (50 nM) of fluorescent pppRVL, resulted in a measurable anisotropy change (**Figure 8**). The data was fitted according to a one site binding model and dissociation constants (K_d) were calculated by least square fitting.

Clearly, the RIG-I WT (full-length) protein showed the highest affinity (**Table 10**). However, most of its binding capacity to pppRVL resided within the RD, as this by itself had an only slightly elevated K_d in comparison to the full-length protein. Mutations in the positively charged groove of RD had a significant impact on pppRVL binding. The K858A mutation almost completely abolished RNA binding. Importantly, the control mutations outside the postulated binding site did not affect the affinity to pppRVL. In summary, these results clearly identified the positively charged cleft within the RIG-RD domain as the binding site for the 5'-triphosphate motif.

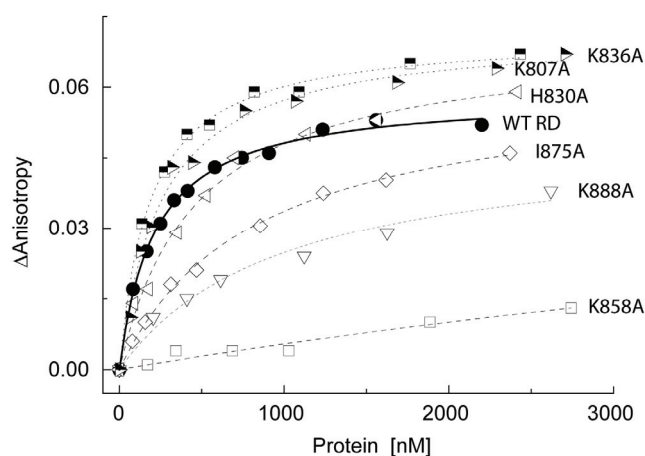


Figure 8: Impact of introduced mutations in RIG-I RD on RNA binding affinity

Fluorescence anisotropy changes (Δ Anisotropy) of fluorescently labeled pppRVL in response to titration with WT RD and the indicated RD mutants, respectively.

Table 10: RIG-I RNA binding affinities

No mutations

RIG-I WT (full-length)	$K_d = 151 \pm 8$ nM
RIG-I RD	$K_d = 217 \pm 18$ nM

Mutations inside the positive charged groove

RD K858A	$K_d > 5,000$ nM
RD K888A	$K_d = 1,003 \pm 180$ nM
RD I875A	$K_d = 1,012 \pm 109$ nM
RD H830A	$K_d = 495 \pm 34$ nM

Control mutations on the convex side

RD D836A	$K_d = 185 \pm 15$ nM
RD K807A	$K_d = 254 \pm 16$ nM

3.2 Functional Dissection of the Individual RIG-I Domains

The RD comprises only a very small part of the RIG-I full length protein. In addition, RIG-I harbors tandem CARDs and a central helicase domain (**Figure 9A**). Importantly, intact RIG-I ATPase activity is crucial for RIG-I mediated antiviral signaling (Gack et al. 2008). To investigate the importance and functional contribution of RD in the context of the full-length

protein, ATPase assays were performed with the complete RIG-I protein and several RIG-I truncation variants lacking either RD and CARDS or both domains.

To this end, several RIG-I truncation mutants were purified according to established protocols (Cui et al. 2008). The RIG-I- Δ RD construct was provided by Sheng Cui. The purified RIG-I constructs (**Figure 9B**) were then used in biochemical ATPase assays, RNA binding studies and single molecule translocation experiments.

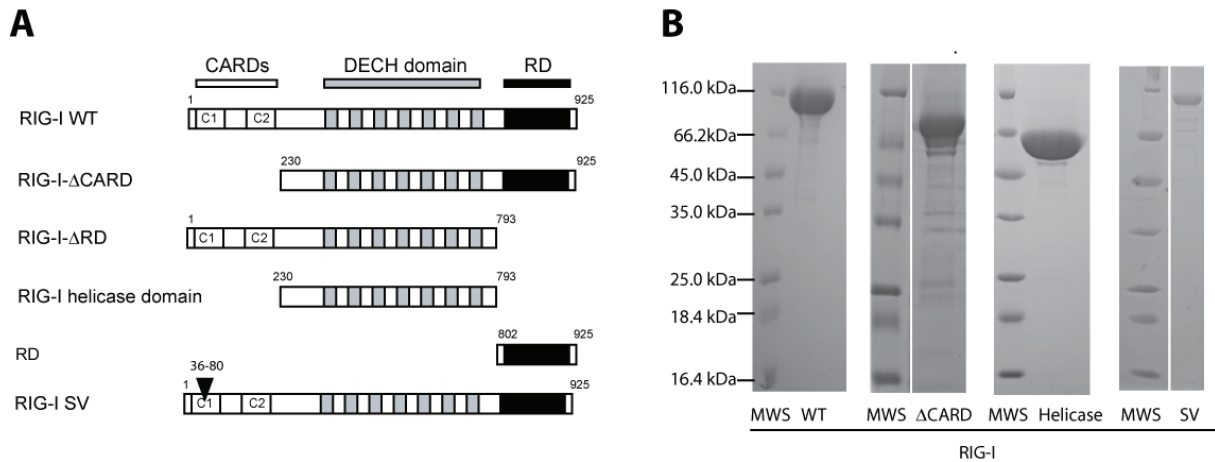


Figure 9: RIG-I constructs used in functional studies.

(A) Schematic representation of constructs used. Domains and domain borders are annotated. RIG-I SV lacks amino acid 36-80 (black triangle) in the first CARD (C1). (B) Coomassie gel showing the final purity of selected constructs and molecular weight standards (MWS) are shown as annotated.

Catalytic efficiencies were determined (k_{cat}/K_M) by measurement of the initial reaction velocity at increasing ATP concentrations. As exemplarily shown for the RIG-I- Δ CARD construct, the reaction mixtures were separated by thin layer chromatography and visualized using trace amounts of γ - 32 P-ATP (**Figure 10**). Two types of RNA were used for stimulation: *in vitro* transcribed pppRVL and synthetic dsRNA without modification at the termini.

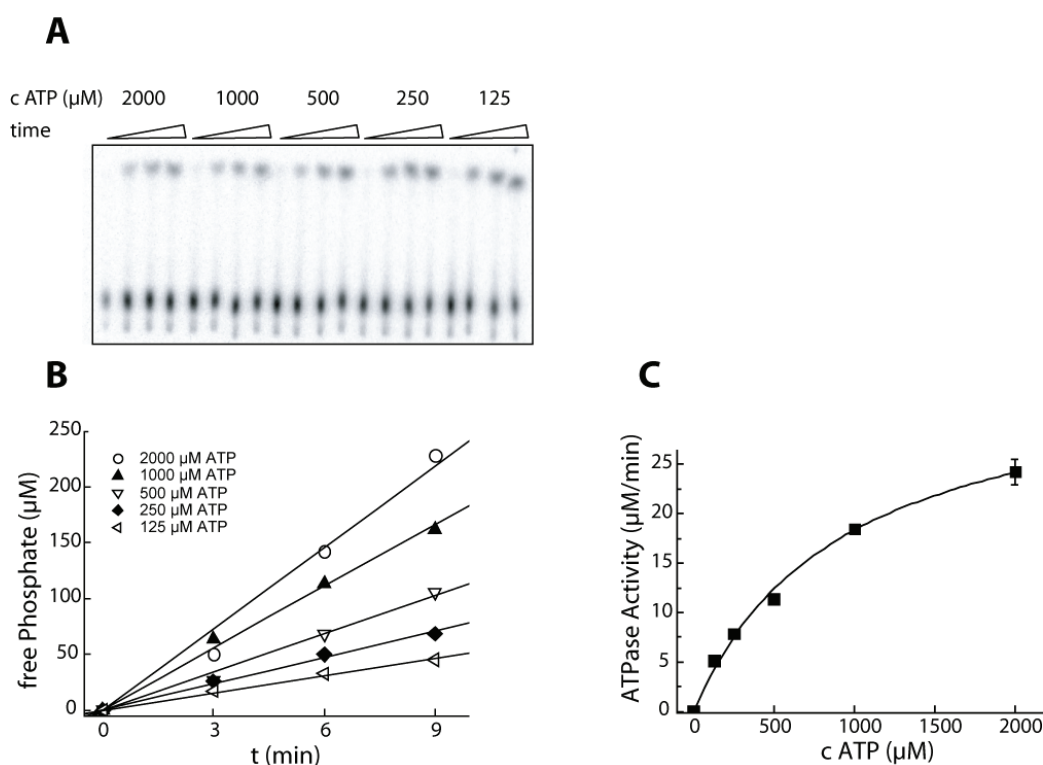


Figure 10: Determination of catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) of the RIG-I- Δ CARD construct.

(A) Thin layer chromatography of ATPase reactions for a representative RIG-I construct (in this case RIG-I- Δ CARD) at 50 μM RVL RNA concentration and decreasing ATP concentrations. (B) ATPase hydrolysis was quantified and initial reaction velocities determined by phosphor imaging. (C) Plotting of these initial velocities against ATP concentration revealed Michaelis-Menten kinetics and allowed determination of the catalytic efficiencies by least square fitting.

A comparison of the catalytic efficiencies revealed that RIG-WT protein was best stimulated by pppRVL. Only very poor ATPase activity was detected with non-phosphorylated dsRNA (**Figure 11A**). The Δ CARD-RIG-I construct showed a similar ATPase activity as the WT protein but, remarkably, had lost its preference for pppRVL with dsRNA also being an equally good stimulator. In comparison to the full length protein, removal of the CARDS seemed to relieve an inhibitory effect on the dsRNA-induced ATPase activity. The isolated helicase domain retained a residual activity, however dsRNA could activate its ATPase activity more efficiently than pppRVL. RIG-I- Δ RD showed no ATPase activity. Therefore, a functional RD seems to be essential for RIG-I catalytic activity.

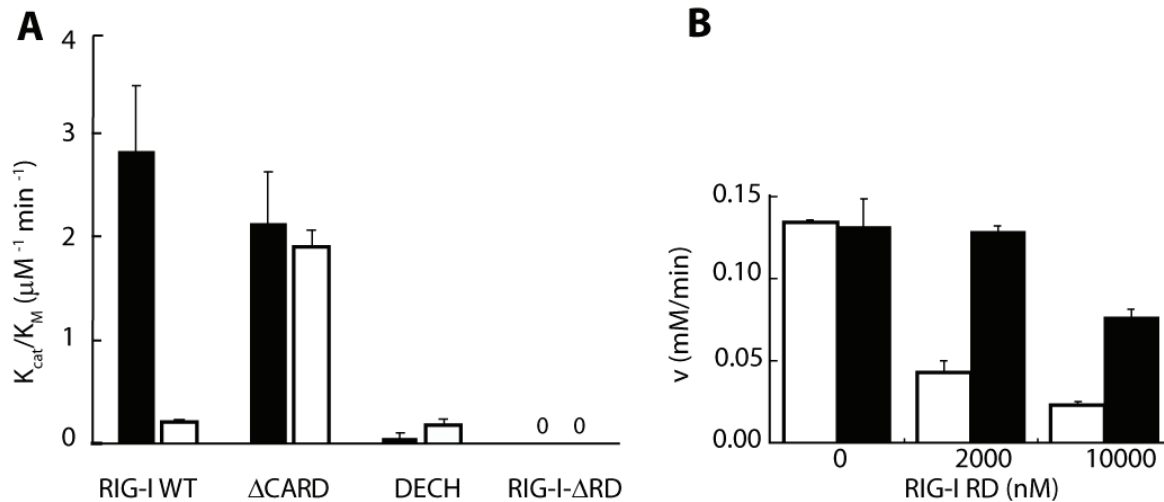


Figure 11: Comparison of ATPase activities of RIG-I variants and dose dependent inhibition by RD.

(A) RIG-I WT and truncation variants were stimulated with pppRVL (black bars) or non-phosphorylated dsRNA (white bars). Error bars represent standard deviations of the non linear regression analysis of catalytic efficiencies (k_{cat}/K_M). (B) Inhibition of the initial ATPase velocity (v) of 20 nM RIG-I- Δ CARD by adding indicated amounts of RD in the presence of 200 nM pppRVL (white bars) or 200 nM dsRNA (black bars). Data shown in A and B represent means and standard deviations (error bars) of three independent experiments.

It has been previously demonstrated that overexpression of RD inhibits RIG-I signaling (Saito et al. 2007). Knowing that RD strongly binds to 5'-triphosphate ends of single stranded RNA, it was hypothesized that the inhibitory effect seen on interferon production might be due to competition of RD with full-length RIG-I for the stimulatory RNA. RD would efficiently mask 5'-triphosphate ends in a way that the full-length protein can no longer be activated. To test this idea, increasing amounts of RD were titrated into a reaction mixture containing Δ CARD-RIG-I (**Figure 11B**). Whereas RIG-I- Δ CARD in the absence of additional RD was equally well stimulated by pppRVL and dsRNA, increasing amounts of RD indeed inhibited pppRVL-stimulated RIG-I ATPase activity in a dose dependent manner. In contrast, dsRNA-induced activation of RIG-I was not significantly affected. Importantly, these results complement the published *in vivo* data and confirm our competition hypothesis. Moreover, these results independently corroborate the conclusion drawn from the direct binding assay (**Figure 8**), that is the potent and specific binding of 5'-triphosphate RNA to the RIG-I RD.

In addition to the functional studies with RIG-I truncation variants, the ATPase activity and RNA binding capacity of the RIG-I splice variant (SV) was also analyzed. This protein carries a deletion within the first CARD (Δ 36-80) and thereby loses its downstream signaling ability due to a failure to interact with the ubiquitin ligase TRIM25 (Gack et al. 2008). Importantly, RIG-I SV acts in a dominant negative manner on the RIG-I-mediated antiviral interferon response. However, the molecular mechanism of inhibition is unclear. One possible

explanation would be the sequestration of viral RNA. However, to be able to compete for viral RNA ligands, the binding ability of RIG-I SV to 5'-triphosphate RNA must still be intact. Another intriguing question is if the amino acid deletion within the first CARD has any influence on the ATPase activity of RIG-I SV. As described above, a similar effect was observed for the RIG-I- Δ CARD construct.

Detailed biochemical analysis showed that indeed RIG-I SV ATPase activity was stimulated by pppRVL, as well as dsRNA in the same order of magnitude as RIG-I WT (**Figure 13A**). Importantly, both proteins show a similar affinity (RIG-I WT $K_d = 246$ nM; RIG-I SV $K_d = 233$ nM) to pppRVL as determined by fluorescence anisotropy (**Figure 13B**). These data suggest that RIG-I SV inhibits interferon signaling by sequestering viral RNA into signaling-incompetent protein complexes. Notably, RIG-I SV retains a preference for pppRVL over dsRNA. This suggests that despite being truncated, the CARDS still exert an inhibitory influence on the ATPase-activation by dsRNA.

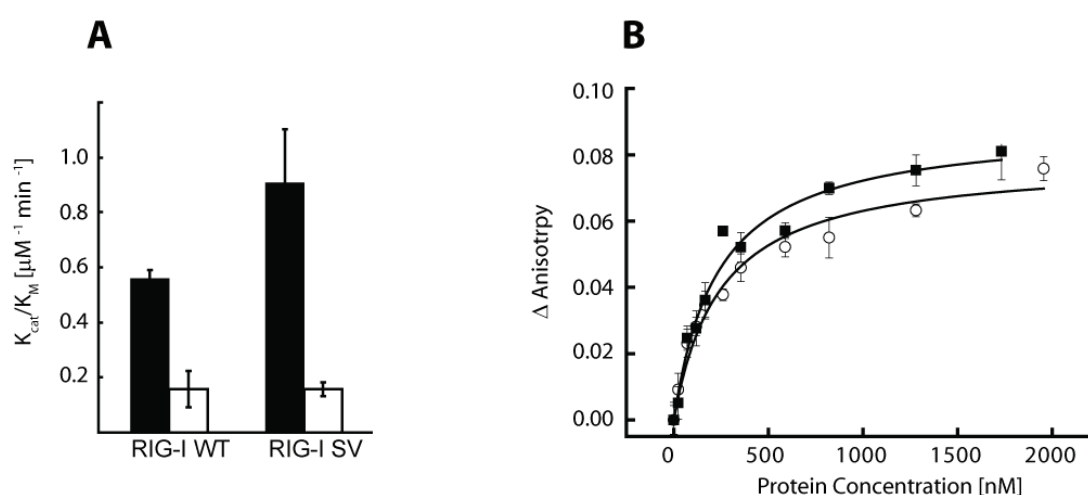


Figure 12: ATPase activity and RNA affinity of the RIG-SV in comparison to RIG-I WT.

(A) RIG-I WT and RIG-I SV were stimulated with pppRVL (black bars) and non-phosphorylated dsRNA (white bars). The ATPase efficiency (k_{cat}/K_M) of RIG-SV is in a similar range as the one of RIG-I WT. (B) Fluorescence anisotropy changes (Δ Anisotropy) of fluorescently labeled pppRVL in response to titration with RIG-I WT (open circles, $K_d = 246 \pm 42$ nM) and RIG-I SV (filled square $K_d = 233 \pm 35$ nM). The deletion in RIG-I SV does not seem to affect RNA binding.

Next, it was investigated if RIG-I, which harbors a helicase motif, was able to translocate along an RNA double-strand in an ATP dependent manner. Since protein dynamics are difficult to follow with classical biochemical methods, RIG-I variants were tested in a setup allowing the tracing of moving single molecules (**Figure 9**). These experiments were performed by the group Prof. Taekjip Ha which was provided with purified RIG-I protein constructs.

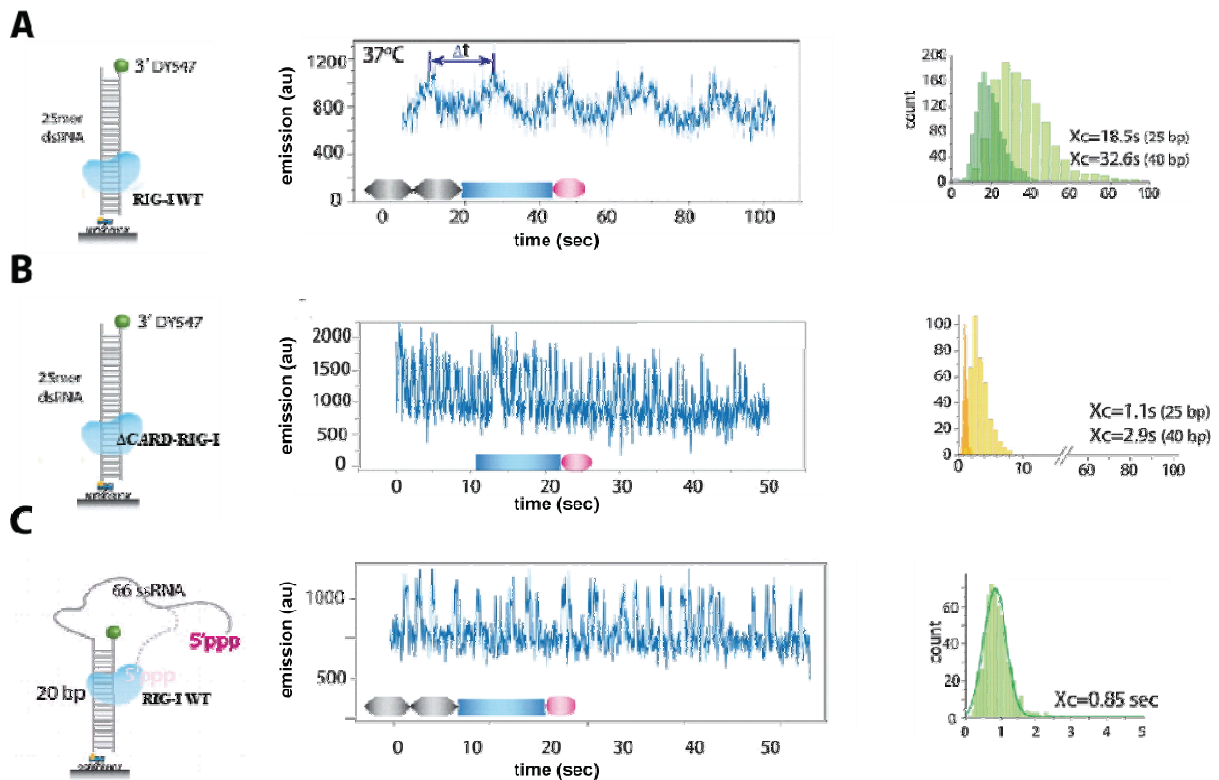


Figure 13: RIG-I shows translocation activity in a single-molecule setup.

(A) A 25mer dsRNA fluorescently labeled with DY547 was immobilized via a biotin modification to a streptavidin-coated matrix (illustrated in the left panel). Addition of RIG-I-WT and ATP leads to a repetitive association with (fluorescence increase) and dissociation from (fluorescence decrease) a single RNA molecule (middle panel). The dwell time (X_c) calculated by Gaussian fitting of all Δt values allows calculation of a translocation velocity ($1/X_c$). (B, C) Similar setups as described in (A) using the Δ CARD-RIG-I construct instead of RIG-WT (B) and a dsRNA with a single stranded extension and a 5'-triphosphate modification (C), respectively. Adapted from (Myong et al. 2009).

RIG-I WT protein showed a robust, ATP-dependent translocation activity on dsRNA (**Figure 13**). Translocation in this context is defined as a movement along the axis of the RNA double strand. By employing a newly developed method, Protein Induced Fluorescence Enhancement (PIFE), protein dynamics could be followed without a fluorescent label on the protein. PIFE relies on the spectral properties of the DY547 dye, which emits fluorescence signals corresponding to the proximity of the unlabeled protein. Two types of immobilized RNA were used for RIG-I stimulation; either an RNA duplex of 25 bp or a 20mer duplex with a single stranded 5'-overhang plus a triphosphate modification (**Figure 13A, C; left panels**). The velocity of translocation was determined by calculating the average time interval between successive intensity peaks (X_c): the shorter the time interval, the higher the enzymatic function. Notably, the translocation activity of RIG-I was much more rapid in the presence of 5'-triphosphate ($x_c = 0.85$ s) in comparison to blunt-ended dsRNA (18.5 s). In accordance with our ATPase studies (**Figure 11**), RIG-I- Δ CARD was again much more active

on unmodified dsRNA ($x_c = 1.1$ s) than RIG-I WT ($x_c = 18.5$ s) (**Figure 13A, B**). This finding further supports a model in which the CARDs exert an inhibitory effect on the helicase (ATPase) domain. Intriguingly, the inhibition by the CARDs could be relieved by the presence of a 5'-triphosphate modification of the RNA ligand (**Figure 13 C**). It is also noteworthy, that RIG-I apparently needs the combination of both molecular patterns to reach full enzymatic activity; a single stranded RNA with 5'-triphosphate ends and dsRNA portions.

3.3 Functional Analysis of MDA5 and Its Interaction with *Paramyxovirus* V-Protein

MDA5 and RIG-I both act as cytosolic receptors for viral RNA. However, they recognize different types of viral RNA. Whereas RIG-I-dependent interferon signaling is specifically induced by 5'-triphosphate RNA and short dsRNA, MDA5 signaling is specifically inducible by poly I:C (Hornung et al. 2006; Kato et al. 2006; Pichlmair et al. 2006). poly I:C is a hybridized complex of a poly inosine (I) and a poly cytosine (C) strand, and is thought to mimic dsRNA. Despite extensive research on RLRs over the past five years, the physiological ligand recognized by MDA5 remains undefined. It had previously been shown that the length of dsRNA might be an important criterion (Saito and Gale 2008).

To identify the physiological MDA5 ligand, a MDA5 ATPase assay in accordance to the RIG-I functional test was established. In this new setting poly I:C was used as the positive control. The utilized protein constructs and their final purity are shown in **Figure 14A** and **B**. Indeed, poly I:C is able to stimulate MDA5 ATPase activity *in vitro* (**Figure 14C**). Interestingly, MDA5- Δ -CARD has a higher ATPase activity than MDA5 FL. This finding is reminiscent of the functional properties of RIG-I where the FL is also less active than the RIG-I- Δ CARD construct when stimulated with dsRNA (chapter 3.2). Therefore, in both cases the CARDs seem to have an inhibitory effect on ATPase activity.

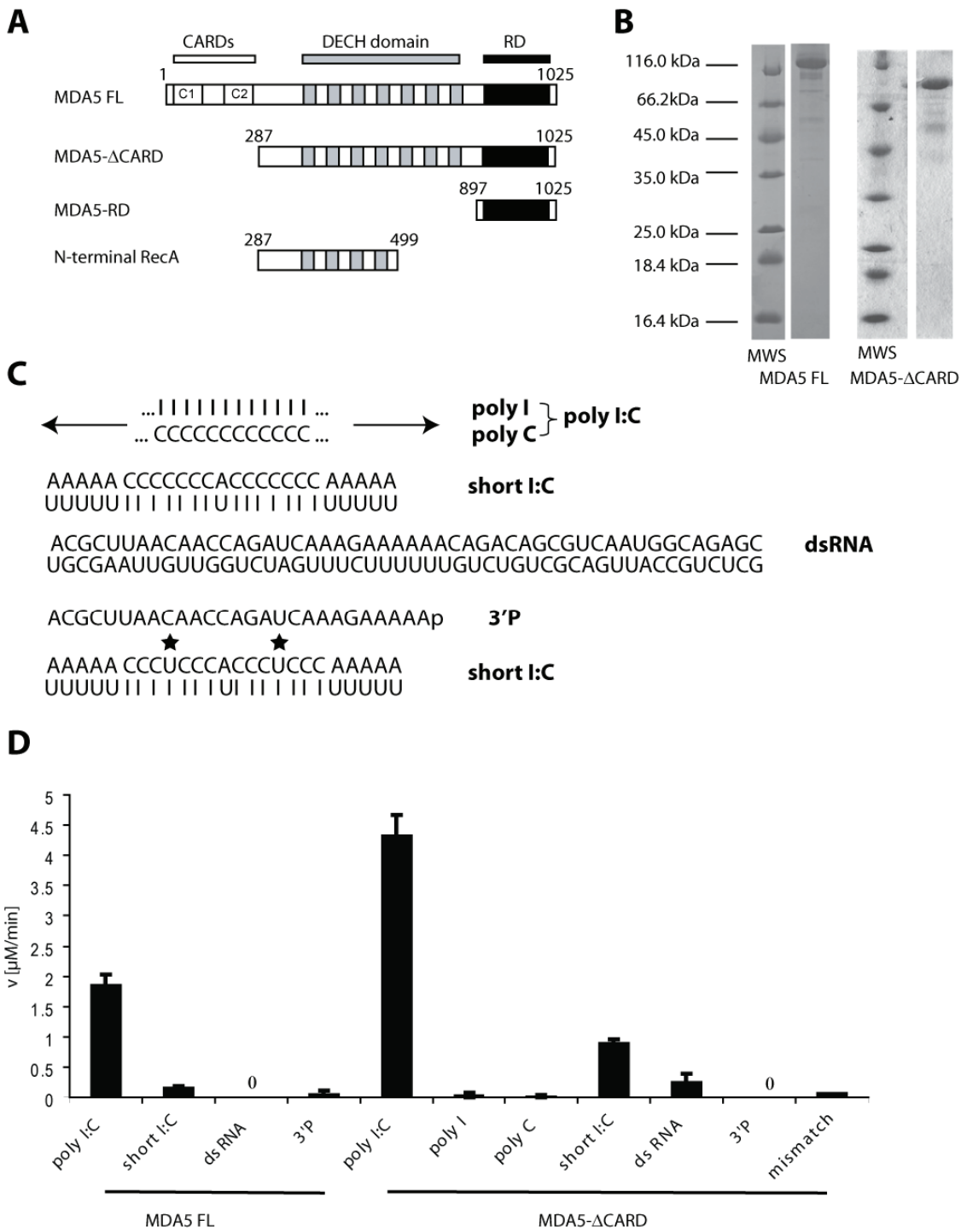


Figure 14: MDA5 ATPase stimulation by various RNA structures.

(A) Domain architecture of MDA5 constructs utilized in functional and structural studies. Domains are colored by the scheme as indicated at top. Terminal amino acid positions are numbered. (B) SDS-PAGE analysis using Coomassie staining of MDA5 FL and MDA5- Δ CARD. Final purified protein as used for ATPase assays with molecular weight standard (MWS) annotated. (C) RNA sequences used to stimulate MDA5 ATPase activity. (D) MDA5 FL and MDA5- Δ CARD (concentration 100 nM) were stimulated by 50 nM of the respective nucleic acid in the presence of 100 μ M ATP. Initial reaction velocities (v) were determined by thin layer chromatography and phosphorimaging as described for **Figure 10**.

Furthermore several potential MDA5 ligands were tested for their ability to activate the MDA5-ATPase (**Figure 14C, D**). Since MDA5- Δ CARD could be purified more easily and showed higher activity in initial tests, this construct was used preferentially during the following analysis. MDA5 activation is absolutely dependent on double stranded oligonucleotides, since single-stranded poly I and poly C alone could not activate MDA5 significantly. Additionally a single stranded RNA with a 3'-monophosphate (3'P), which had been described as a potent RIG-I activator (Malathi et al. 2007), also failed to induce MDA5 activity. Notably, the length of the double strand seems to be an important criterion, since a chemically synthesized oligonucleotide with a high I:C content and a length of 25 bp (short I:C) was less stimulatory than poly I:C (a heterogeneous mixture of polynucleotides with a length >2000 bp). Interestingly, the sequence and the resulting structural features of the RNA seem to play an important role in its capability to stimulate MDA5. Synthetic dsRNA built from the rabies virus leader sequence paired to its complementary strand (length 50 base pairs) was less stimulatory than short I:C. Furthermore, when introducing an I:U mismatch into the short I:C double stranded oligonucleotide, its ability to induce MDA5 ATPase activity was significantly reduced.

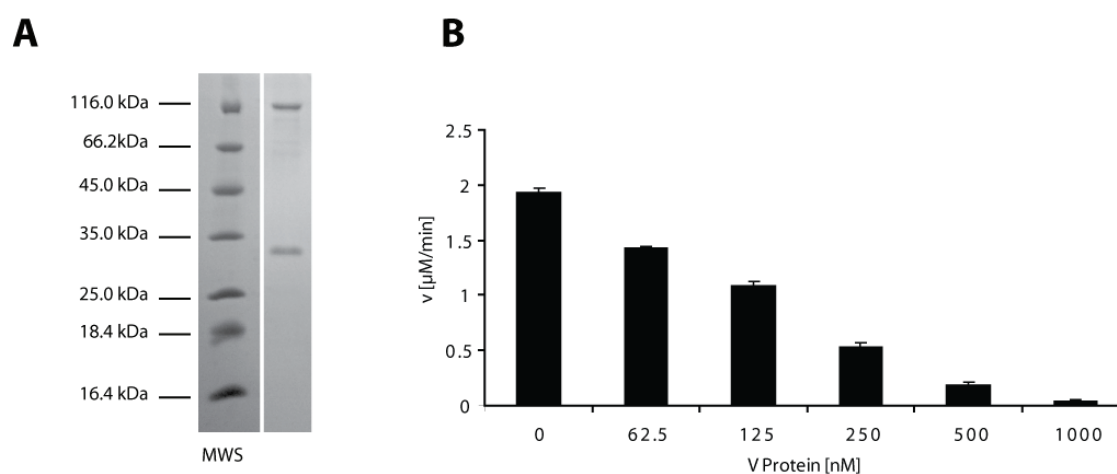


Figure 15: MDA5 ATPase activity is inhibited by increasing amounts of V-protein.

(A) MDA5 and V-protein complex co-purified by gelfiltration chromatography. SDS-Page analysis and Coomassie staining of MDA5 FL (~120 kDa) and an N-terminally truncated version of Measles Virus V-protein (residues 49 – 299, ~30 kDa). (B) MDA5 ATPase activity is inhibited by V-protein in a dose-dependent manner. MDA5- \square CARD (100 nM) was stimulated by 50 nM poly I:C in the presence of 100 μ M ATP and increasing amounts of Measles Virus V-protein. For each reaction initial ATPase reaction velocities (v) were determined.

Paramyxovirus V-proteins interfere with MDA5-dependent signaling by binding to MDA5 via a conserved cysteine-rich C-terminus, thereby inhibiting its function (Andrejeva et al. 2004; Childs et al. 2007). It was tested whether the detrimental interaction found *in vivo* could also be reconstituted *in vitro*. Indeed, when co-expressed, full length as well as N-terminally

truncated Measles Virus and Simian Virus 5 (SV5) V-protein with a MDA5 FL could be co-purified (**Figure 15A**).

The structural and mechanistic details of the MDA5 V-protein interaction are still poorly understood, including which domains within MDA5 are targeted, and whether the ATPase activity, essential for downstream signaling is affected by V-protein binding. These questions were addressed by performing the previously described MDA5 ATPase assay (**Figure 15B**). Increasing amounts of purified Measles Virus V-protein (residues 49 – 299) were titrated to poly I:C stimulated MDA5- Δ CARD. Increasing amounts of V-protein were found to inhibit the ATPase reaction in a dose-dependent manner. At a saturating concentration of V-protein (1 μ M), less than 5% residual MDA5 activity was observed.

Since V-protein seems to interfere directly with MDA5 ATPase activity it is speculated that its inhibition is due to a direct targeting of the helicase (DECH box) motifs. In a primary unbiased approach, the regions of MDA5 that interact with V-protein were mapped using a peptide interaction scan. Overlapping, MDA5-derived peptides of 12 amino acids length were probed for interaction with myc-tagged SV5 V-protein (**Figure 16A**). The peptide array has the advantage of covering the entire polypeptide in fine resolution. However, only linear epitopes can be recognized, while epitopes present only in the tertiary structure cannot be reconstituted on the array. Additionally, regions of the protein that are normally buried in the hydrophobic core might be exposed and cause false positive interactions. Therefore protein interactions indicated by this assay need to be confirmed by further independent experiments. Nevertheless, the peptide screen yielded interesting results. Interacting peptides were found within both the N-terminal, as well as within the C-terminal RecA fold of the MDA5 helicase domain, suggesting that V-protein indeed binds to the DECH box domain (**Figure 16B**). Surprisingly additionally interacting peptides were also found within the first CARD as well as within the C-terminal RD. However, as these results could not be confirmed by *in vitro* interaction studies with purified protein components (data not shown), they probably represent experimental artifacts.

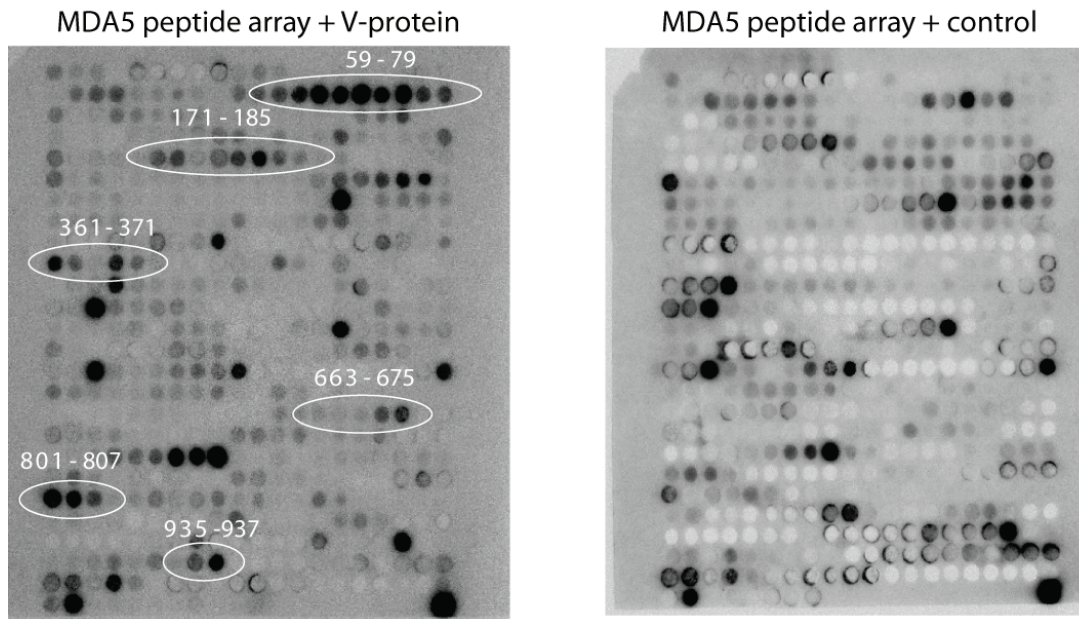
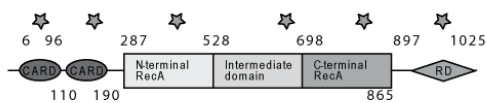
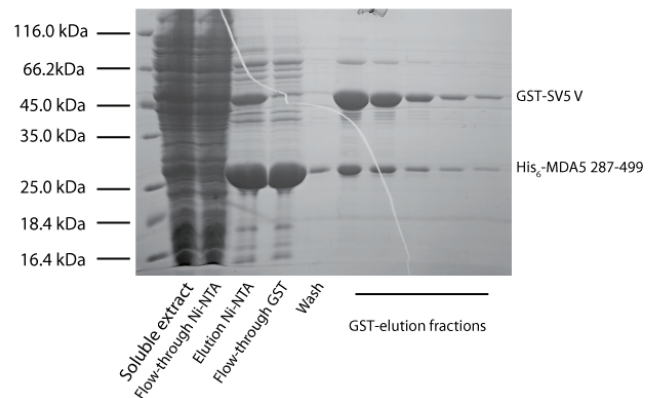
A**B****C**

Figure 16: A Peptide SPOT assay reveals multiple V-protein MDA5 interaction sites.

(A) A MDA5 peptide-array was probed with V-protein or a negative control protein (myc-GST). Non-specific interactions in the control blot (right panel) allow the identification of specific MDA5 / V-protein interactions (MDA5 amino acid positions annotated). (B) The interaction sites (grey stars) were mapped on the MDA5 sequence. (C) SDS-PAGE analysis and Coomassie staining of co-purified MDA5 V-protein complex. GST tagged V-protein (GST-SV5V) remains bound to His tagged MDA5 N-terminal RecA domain (His₆-MDA5 287-499) after a tandem affinity purification with Ni-NTA followed by purification with GST-sepharose. Molecular weight standard (MWS) with respective sizes and respective fractions during purification are annotated.

By co-expressing a series of truncation variants of the MDA5 helicase domain the MDA5 V-protein interaction site was further narrowed to the N-terminal Rec A domain of MDA5, which harbors the critical helicase motifs I, Ia, II, III (**Figure 14A**). The N-terminal RecA fragment (MDA5 297-499) co-purified with V-protein during dual affinity purification (**Figure 16C**) and was therefore sufficient for V-protein binding. Although a stable homogenous complex could

be purified, further structural analysis by X-ray crystallography was hindered by the relatively high instability of the V-protein constructs.

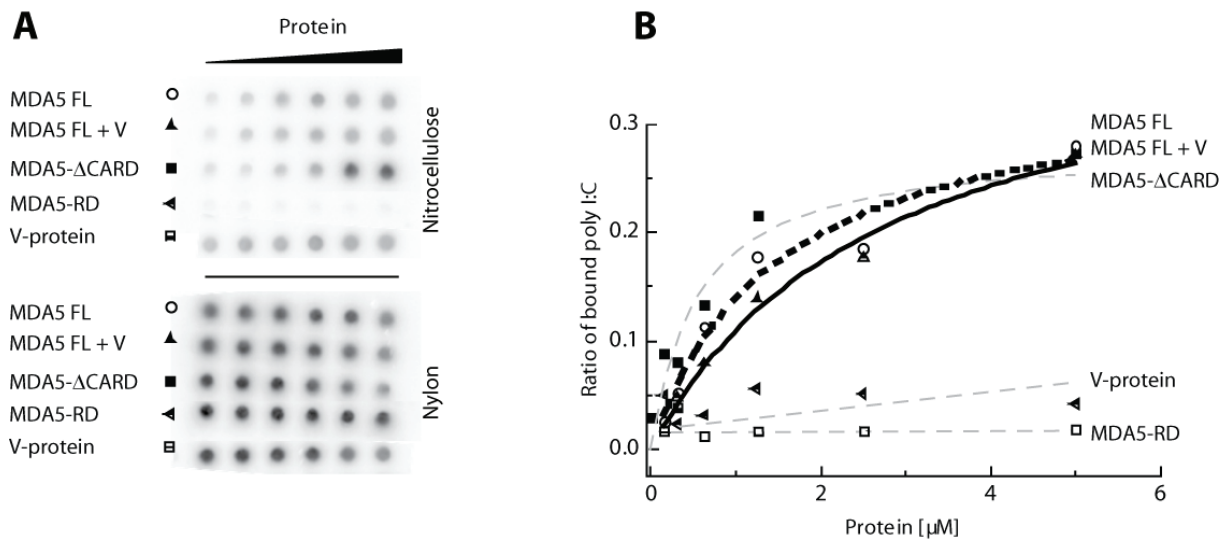


Figure 17: V-protein does not influence binding affinity of poly I:C to MDA5.

(A) Affinities of ^{32}P -labeled poly I:C were determined by dot blot filter binding assays. Increasing protein concentrations were incubated with 50 μM poly I:C. MDA5 FL was inhibited with an excess of 5 μM V-protein. (B) The blots were quantified by phosphorimaging. V-protein neither sequesters poly I:C directly (open squares), nor does it inhibit binding of MDA5 to poly I:C (open circles and dashed black line MDA5 FL alone; up facing triangles and solid black line MDA5 FL + 5 μM V-protein).

To learn more about the mechanism of inhibition, the binding of V-protein to MDA5 was tested to see if access of RNA to the helicase domain could be blocked. The affinity of poly I:C to MDA5 was measured in the presence and absence of an inhibiting amount of V-protein by utilizing a dot blot filter binding assay (**Figure 17**). Poly I:C binding to MDA5 was not affected by the addition of V-protein. Importantly, V-protein alone was incapable of binding to poly I:C thereby excluding the possibility of inhibition by substrate sequestration. Thus, competition with nucleic acid binding is not the mechanism of inhibition of ATPase activity of MDA5 by V-proteins. It is also of note that the binding affinity to poly I:C seems to primarily reside within the helicase domain, since deletion of the CARD domain does not reduce poly I:C affinity and the isolated RD shows only very low binding.

3.4 Structural Studies on the MDA5 Regulatory Domain with Small Angle X-Ray Scattering and Homology Modeling

Our laboratory successfully determined the X-ray structures of the C-terminal RDs of both RIG-I and LGP2 (Cui et al. 2008; Pippig et al. 2009). Clearly, a structural comparison between these structures and the MDA5 RD would be highly informative. However, despite

extensive effort, MDA5 RD resisted crystallization. As an alternative to X-ray crystallography, analysis by small angle X-ray scattering (SAXS) in combination with molecular modeling can provide structural envelope, to allow comparison to the MDA5, RIG-I and LGP2 RDs.

To structurally characterize MDA5 RD a series of MDA5 constructs were cloned and purified from three different species (human, mouse and chicken). **Figure 18** demonstrates the final purity of the human construct 897-1025.

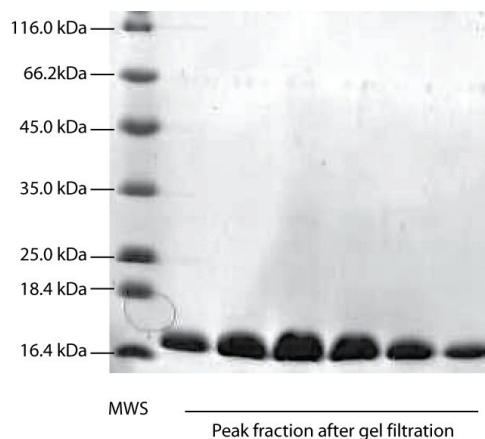


Figure 18: Purification of MDA5 RD.

SDS-PAGE analysis of the MDA5 construct 897-1025 after final gel filtration purification step. Molecular weight standard (MWS) with respective sizes and respective fractions during purification are annotated.

This construct was subsequently analyzed in solution by small angle X-ray scattering (SAXS). The measurements were performed at three different protein concentrations (2, 5 and 10 mg/ml) and the resulting spectra merged after appropriate scaling using the data analysis software package ATSAS (Konarev et al. 2006). After ten individual rounds of *ab initio* modeling with GASBORp the models were manually oriented and averaged resulting in the final envelope shown in **Figure 19B**. SAXS analysis of the RIG-I RD was performed by Sheng Cui and the data provided for a comparison with MDA5 RD data.

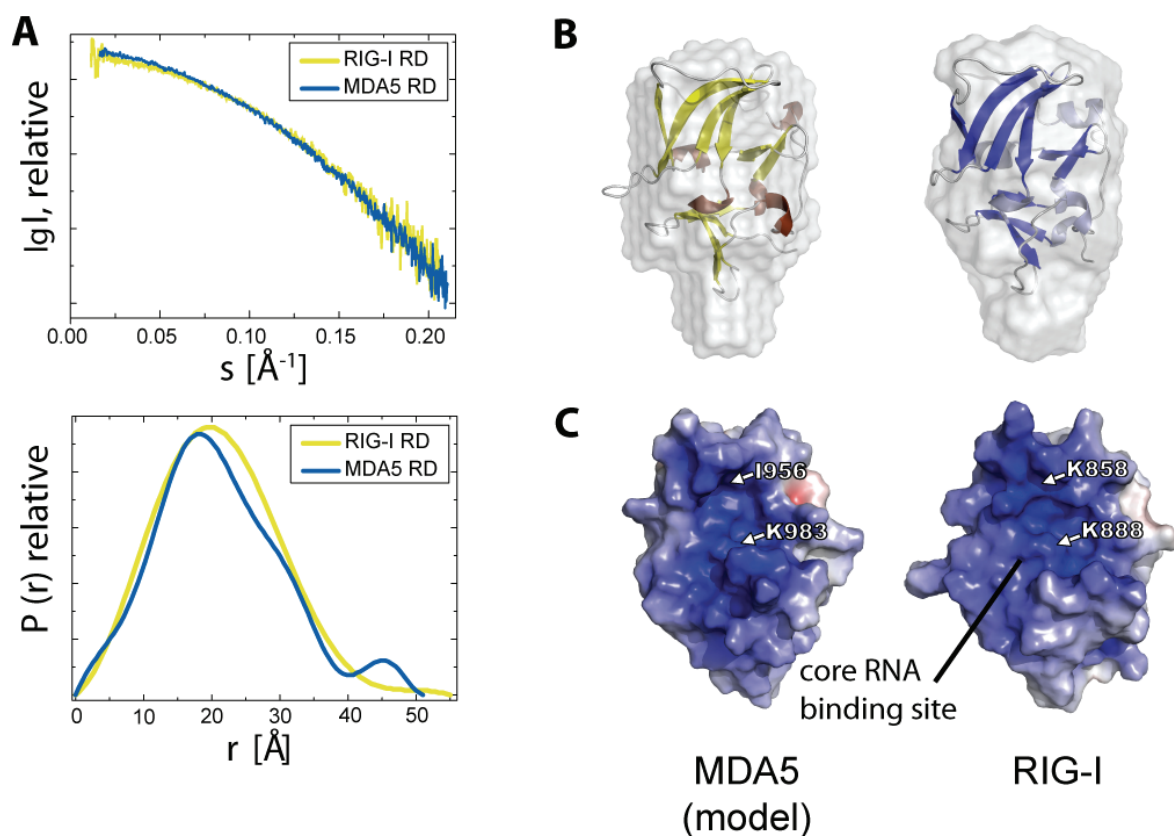


Figure 19: Small angle X-ray scattering analysis and molecular modeling of MDA5 RD.

(A) Primary scattering curves (upper panel) and calculated pair distribution functions (lower panel) of MDA5 and RIG-I RD (colored as indicated). (B) *Ab initio* molecular envelopes (faint gray; RIG-I RD right panel; MDA5 RD left panel) calculated from the primary data as shown in (A) and superimposed with the RIG-I RD crystal structure (yellow) and an MDA5 model (blue). The theoretical model for MDA5 RD matches the experimentally determined SAXS envelope. (C) Comparison of the electrostatic surface potential of MDA5 and RIG-I RD. Both domains harbor a similar positively charged patch, however residues which have been shown to be essential for the RNA-binding capability of RIG-I are not conserved in MDA5 (e.g. K858^{RIG-I} in comparison to I956^{MDA5}).

As already obvious from the primary scattering curves and the calculated pair distribution function (**Figure 19A**) the solution structures of RIG-I RD and MDA5 RD are very similar in dimension. Accordingly, the obtained radii of gyration (R_G) are almost identical ($R_G^{\text{RIG-I-RD}} = 16.2 \text{ \AA}$ and $R_G^{\text{MDA5-RD}} = 16.1 \text{ \AA}$). Furthermore, the calculated envelope of MDA5 correlates very well with the geometry of the RIG-I RD SAXS model and the corresponding crystal structure (**Figure 19B**; left panel). The high conservation of hydrophobic core regions and important structural motifs, such as zinc-coordinating cysteines (**Figure 20**) is also very conspicuous. Taken together these observations suggest that the RIG-I and MDA5 RDs possess an overall similar structure. To obtain deeper molecular insights into the MDA5 RD structure, the program MODELLER was employed to calculate an initial model of the MDA5-RD structure. For this, a sequence alignment of MDA5 RD with the respective domains of RIG-I and LGP2 plus the RIG-I and LGP2 RD crystal structures were used as input. It is likely that the model will differ in precise orientation of flexible loop regions from a respective high

resolution structure. Nevertheless, the calculated model fit the experimentally determined SAXS envelope well (**Figure 19B**; right panel). Interestingly, MDA5 RD shows a similar, positively charged patch as was found in RIG-I RD (**Figure 19C**). This is surprising, because in contrast to RIG-I RD and LGP2 RD, the binding affinity of MDA5 RD to RNA is very low. No binding to poly I:C (**Figure 17**) or to pppRVL could be detected (Cui et al. 2008). As expected from this observation, the critical residues mediating the RNA interaction in RIG-I are not conserved in MDA5 (i.e. K858^{RIG-I} in comparison to I956^{MDA5}, **Figure 19C**).

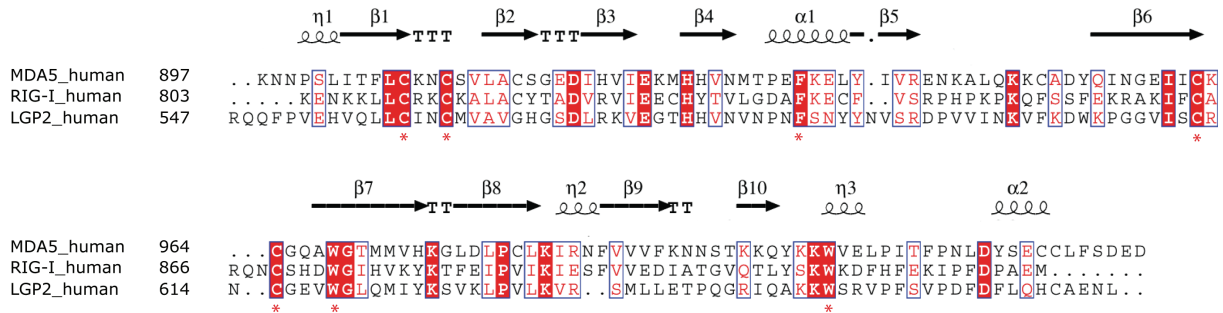


Figure 20: Structure-based sequence alignment of RLR-RDs.

Conserved residues in MDA5, RIG-I and LGP2 RDs are highlighted in red boxes and red letters. The secondary structure for MDA5 as predicted by MODELLER is depicted on top. Residues important for folding such as the zinc-coordinating cysteines and hydrophobic core residues are marked by a red asterisk.

4 Discussion

4.1 The Positively Charged Patch within RIG-I RD is the Recognition Site for 5'-Triphosphate RNA

The C-terminal RD domain of RIG-I is key to the recognition of viral PAMPs. Through structure-guided mutagenesis, it was shown that the positively charged patch, identified in the RIG-I RD crystal structure, is indeed the 5'-triphosphate recognition site in RIG-I. Mutating of several residues within the binding pocket severely reduce the binding affinity to pppRVL. An especially drastic effect is seen when K858 is mutated to alanine, indicating that this residue is central for RNA binding. This particular lysine is situated in the center of the positively charged binding pocket and its location is therefore well suited for phosphate interaction. Interestingly, in LGP2 and MDA5 this side chain is not conserved, being either a proline and or an isoleucine (**Figure 20**), which might explain the observed differences in ligand specificity between the three helicases. RIG-I RD specifically binds to pppRVL and LGP2 prefers dsRNA (Cui et al. 2008; Pippig et al. 2009), whereas MDA5 shows a comparably low affinity to all kinds of RNA tested.

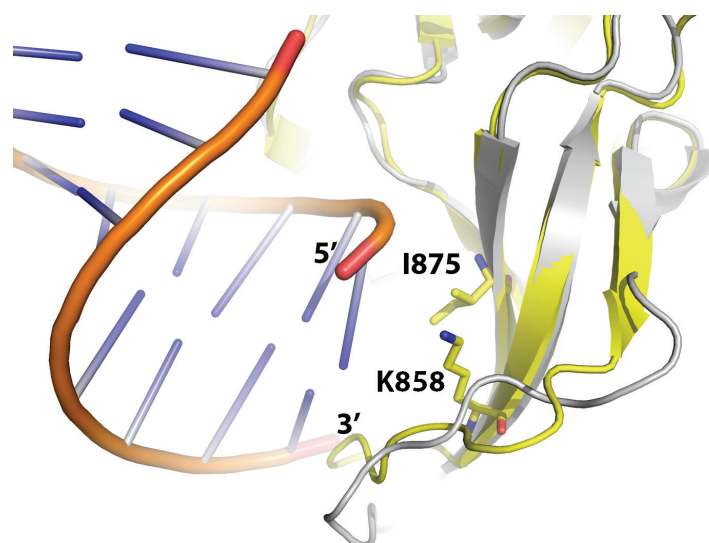


Figure 21: Overlay of the LGP2 RD RNA complex with Rig-I RD.

Superposition (ribbon presentation) of the LGP2 RD (in grey) plus RNA (orange, red and blue) and the RIG-I RD (in yellow). Crucial residues for recognition of the 5'-triphosphate motif within RIG-I are shown as sticks with annotated site chains. The termini of the RNA oligonucleotide are labeled accordingly.

A recently reported structure of LGP2 in complex with a short dsRNA oligonucleotide (Li et al. 2009) reveals that LGP2 RD specifically recognizes RNA termini, preferentially the 5' ends. Interestingly, when superimposing the RIG-I RD structure with the LGP2 RNA complex, the 5' end recognition site within LGP2 co-localizes with the predicted triphosphate recognition site

of RIG-RD. Crucial residues like K858 and I875 are in direct proximity to the overlaid RNA structure, and well positioned to coordinate a triphosphate extension. Although the exact molecular details of the RIG-I RD RNA interaction can only be inferred from an experimentally derived co-complex structure, it seems that recognition of nucleic acids by RIG-I and LGP2 follows common principles. In both cases, small modifications of the RD scaffold allow for specific recognition of RNA-end structures.

4.2 RIG-I Integrates Two Pathogen-Associated Molecular Patterns

Functional analysis of RIG-I revealed that dsRNA as well as single stranded RNA with 5'-triphosphate extensions can induce enzyme activation (**Figure 11**). These two distinct viral patterns are also sensed by two separate domains within RIG-I. Whereas the RD seems to be crucial for recognizing 5'-triphosphate end structures, the helicase domain binds and requires dsRNA to activate its ATPase-mediated translocation activity. *In vivo*, both patterns are most likely present on one RNA and are probably also required for a full antiviral response. This hypothesis seems challenged by the finding that *ex vivo*, single stranded pppRVL alone is a very potent RIG-I activator. However, it is very likely that pppRVL is not strictly single-stranded but also forms double-stranded RNA structures. Bioinformatic analysis of the rabies virus leader sequence revealed an intrinsic tendency of pppRVL to form double-stranded hairpins (Schmidt et al. 2009). Furthermore, there is accumulating experimental evidence that the active species in T7 polymerase transcripts, such as pppRVL, is actually a byproduct resulting from template switching of the polymerase. The resulting RNA would have an extended intrinsically complementary sequence, which would trigger hairpin formation (Schmidt et al. 2009). Therefore, the 5'-triphosphate motif of such an RNA would be recognized by RD whereas the double stranded parts would simultaneously and cooperatively activate the helicase domain. In line with this model, the molecular dimensions of the RIG-I RD allow for recognition of the 5'-triphosphate end plus a maximum of three additional nucleotides only. It is therefore highly likely that, in the context of the full-length protein, 3' protruding nucleotides forming double-stranded structures will reach the helicase domain to induce full activation of the protein.

4.3 CARDS Play a Dual Regulatory Role

The RIG-I CARDS seem to have two opposing functions: they act (1) as a negative regulator of the helicase (ATPase) domain, potentially by sterically blocking access to the dsRNA binding sites in the helicase (DECH box) motif and (2) as the positive mediator of

downstream signaling by facilitating CARD / CARD interaction with the adaptor molecule IPS1. Deletion of the complete CARDS results in the loss of preference for 5'-triphosphate RNA so that the truncated RIG-I is now equally-well stimulated by dsRNA bearing unmodified ends. In comparison, the RIG-I splice variant, which harbors a deletion in the N-terminal CARD, loses its capability to interact with IPS1 but retains its auto-inhibition of the ATPase activity. The observed up-regulation of RIG-I SV expression after induction of anti-viral signaling therefore may represent a potent negative feedback mechanism. Signaling defective RIG-I splice variant would then sequester viral RNA into inactive hetero-complexes, thereby shutting off again the anti-viral signaling cascade, once sufficient activation has occurred.

4.4 Proposed Model for RIG-I Activation

In summary, the data presented here suggest a step-wise model of RIG-I activation, which can be described as the following chain of events (**Figure 22**). In the inactive form, RIG-I is monomeric and arranged in a conformation in which the CARDS are buried within the helicase domain, blocking the access of and subsequent stimulation by dsRNA species. This is demonstrated by the observed inhibitory effect of the CARDS on RIG-I ATPase activity when stimulated with dsRNA. Upon binding of RNA bearing a 5'-triphosphate modification, RD mediates dimerization of the protein (Cui et al. 2008). Furthermore, via an extensive conformational change of the CARDS, the helicase domain becomes exposed and is now free to interact with dsRNA. Activated RIG-I now presumably translocates on dsRNA in an ATP dependent manner (Myong et al. 2009), thereby exposing the CARDS for downstream interaction with IPS-1. What might be the functional relevance of the RIG-I translocation activity? First, by moving on dsRNA RIG-I might effectively displace viral proteins, thereby interfering with viral replication and de-masking viral patterns. Furthermore, if RIG-I in translocation mode actually represents the actively signaling form of the protein, then signal strength would be limited by the time spent in translocation mode. This might explain why RIG-I and MDA5 differ in their ability to signal in response to long double-stranded RNA.

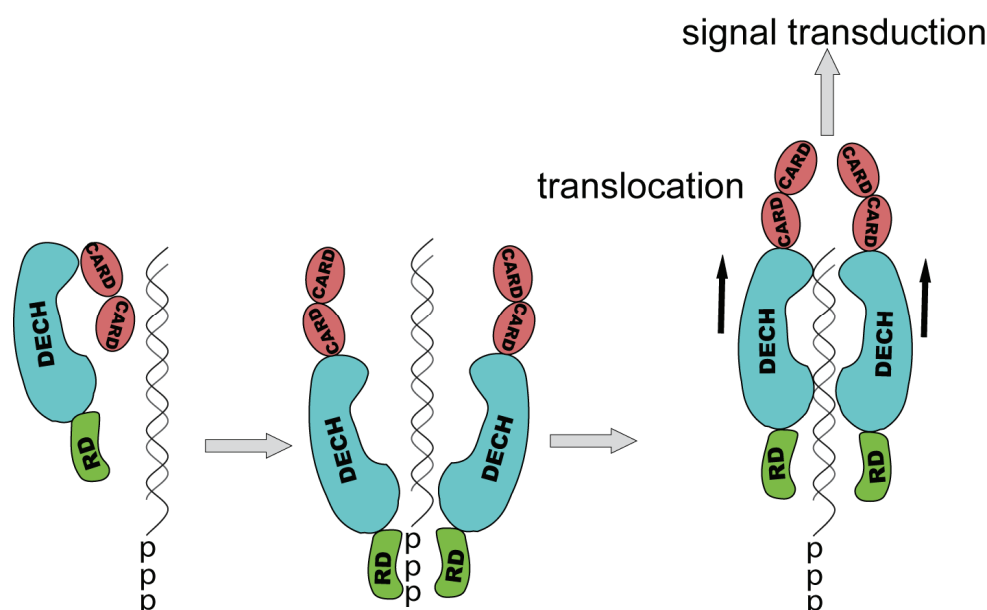


Figure 22: Proposed model for RIG-I activation.

Schematic illustration of events leading to RIG-I activation. In the inactive state, the RIG-I CARDs block access to the helicase domain (DECH). Binding of the 5'-triphosphate motif to the RD induces dimerization. Subsequently, the CARDs undergo a large conformational change, exposing themselves to interact with downstream signaling factors. In addition, the helicase (DECH) domain mediates ATP-dependent translocation on dsRNA parts.

4.5 Hypothetical Modes of Viral Recognition by MDA5

Since poly I:C is sufficient for stimulation *in vitro*, the physiological ligand recognized by MDA5 is most likely only composed of nucleic acids as well. However, so far no defined molecular (RNA) epitope could be found that is responsible for MDA5 induction. The here conducted functional ATPase assays demonstrated that a double stranded oligonucleotide is needed to induce MDA5 ATPase activity. Most likely, the length of the double strand is an important criterion, as has also been implicated in an independent study (Saito and Gale 2008). Recently, there has been accumulating evidence that MDA5 might not function alone but rather in concert with LGP2 (Venkataraman et al. 2007; Pippig et al. 2009). If this holds true, LGP2 might augment MDA5 dependent signaling, which would explain the difficulties in stimulating MDA5 on its own. However, an LGP2 / MDA5 complex could not be reconstituted *in vitro* and addition of LGP2 had no synergistic effect on MDA5 ATPase activity (data not shown). In a physiological scenario, LGP2 and MDA5 might interact in a rather transient manner, although such interactions would be difficult to trap with biochemical methods. Furthermore, as yet unidentified factors might bridge a functional LGP2 / MDA5 complex and might explain why the individual protein components fail to interact directly *in vitro*.

4.6 V-Protein Interferes with Helicase Activity of MDA5 and Thereby Inhibits Signaling

Mapping the interaction of MDA5 and *Paramyxovirus* V-protein by peptide SPOT analysis as well as co-expression of truncation variants revealed a stable complex between the N-terminal RecA domain of MDA5 and V-protein. It was shown that sequestration of viral RNA by V-protein is not the mechanism of MDA5 inhibition. Rather, RNA is still able to bind to a MDA5-V-protein complex. Therefore V-protein binding most likely freezes MDA5 in a state which is incapable of ATP hydrolysis. Since integrity of the ATPase is necessary for downstream signaling the antiviral response is thereby severely attenuated. A model for the inhibitory role of V-protein is depicted in **Figure 23**.

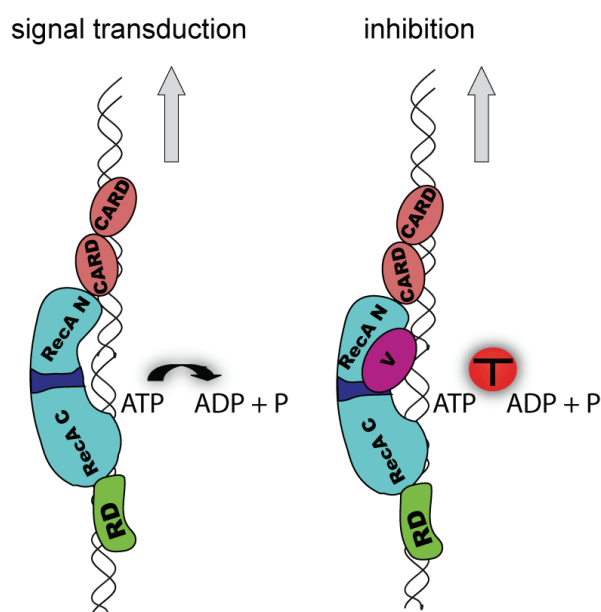


Figure 23: Model for MDA5 inhibition by V-protein.

Binding of V-protein to the N-terminal RecA domain (RecA N) prevents ATPase activity in the presence of viral RNA patterns (long dsRNA). The V-protein MDA5 complex is incompetent of antiviral signaling.

4.7 MDA5 RD Structurally Resembles RIG-I RD

The structural data reported here suggest that the RDs of MDA5 and RIG-I are closely related in fold and share a common RNA binding core. Although both domains harbor a similar, positively charged patch, that is ideally suited for RNA recognition, MDA5 has a substantially lower affinity for RNA ligands. This is most likely due to subtle differences in the key side chains forming the binding pocket. For instance, residue K858^{RIG-I}, which has been shown to be essential for RNA recognition in RIG-I is not conserved in MDA5 (**Figure 20**). During the course of this thesis, a crystal structure of MDA5 RD was published in the protein data bank (Protein Data Bank Accession Code 3GA3). By comparing the experimentally

derived structure with the theoretically calculated model it is now possible to evaluate the reliability of the structure prediction performed in this work (**Figure 24**).

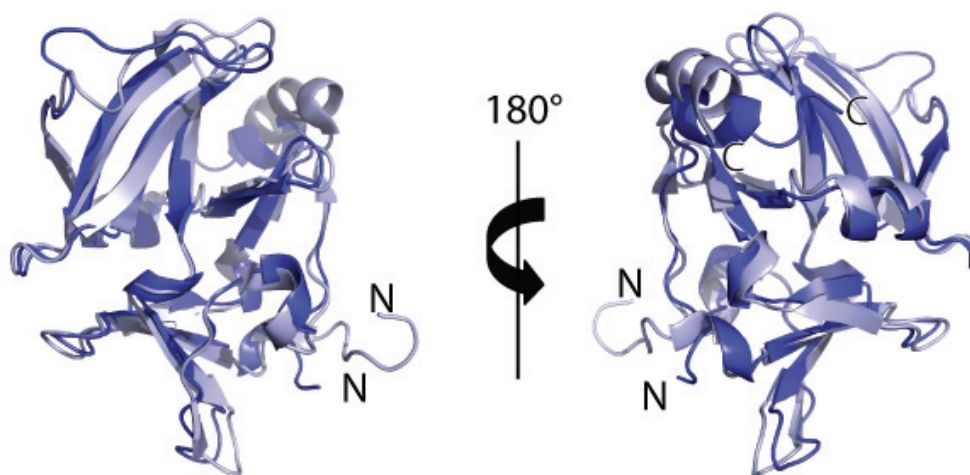


Figure 24: Comparison of the theoretical MDA5 RD model with the experimentally derived crystal structure.

Overall, the MDA5 RD model (dark blue ribbon) and the MDA5 RD crystal structure (light blue ribbon) superimpose well on each other. Slight aberrations are visible in flexible loop regions and the C-terminal α -helix.

Overall, the predicted MDA5 RD structure fits the experimentally determined crystal structure very well. Not surprisingly, the flexible loop regions and the C-terminus, which was artificially modified (His-tag) in the crystal structure, vary in conformation. Indeed, MDA5 harbors a similar positively charged patch as has been shown for RIG-I. This successful example of structural prediction clearly illustrates the power of current homology structural modeling.

5 Summary

The cytosolic helicases RIG-I and MDA5 are primary sensors for viral RNA during infection. Although their overall role as key players in the antiviral response and the induced signaling pathways have been elucidated in great detail over the past years, a structural and functional understanding of virus recognition by these sensors is missing.

On the basis of an X-ray structure of RIG-I RD the 5'-triphosphate interaction site could be mapped to a previously identified positively charged groove, employing structure-guided mutants in RNA binding assays. Structural modeling of the homologous RD of MDA5 implicated that it possesses an overall similar fold and a positively charged binding groove. However, crucial residues for RNA binding in RIG-I RD are not conserved in MDA5, explaining its low affinity to RNA. In the context of the full length protein, RIG-I RD has an important regulatory role on the RIG-I ATPase domain, since its removal drastically impaired ATPase activity. In contrast, the CARDs seem to negatively regulate RIG-I ATPase activity. Based on a comparison of enzymatic activities of several RIG-I truncation variants, a model was postulated in which the CARDs block access of RNA to the helicase motif in the inactive state of RIG-I. Upon recognition of the RNA 5'-triphosphate motif by RD, RIG-I undergoes a conformational change which exposes the CARDs, allowing downstream signaling to IPS-1 as well as binding of RNA to the helicase motif. A collaborative study revealed that RIG-I actively translocates on dsRNA. Furthermore, these experiments emphasized that RIG-I acts as a receptor which integrates a binary molecular pattern. Besides the 5'-triphosphate end structure, which is recognized by RD, double stranded RNA parts bind to the helicase domain and induce ATP-dependent translocation.

In contrast to RIG-I, the molecular patterns which lead to MDA5-dependent anti-viral signaling are not sufficiently understood. This work has shown that the dsRNA-mimic poly I:C is a potent activator of MDA5 ATPase activity *in vitro*, in analogy to its stimulatory effect in cell culture experiments. Interestingly, ATPase activity is inhibited by V-protein, which stably binds to the N-terminal RecA domain of the MDA5 helicase domain. The V-protein mediated inhibitory effect seems to be caused by its interference with ATP hydrolysis directly, rather than RNA sequestering or preventing MDA5 RNA interactions.

In summary, the results presented here broaden our understanding of virus sensing and pinpoint several intriguing avenues for future research.

Part II

Modulating Spectral Properties of the Green Fluorescent Protein with Nanobodies

6 Introduction

6.1 Green Fluorescent Protein: From Initial Discovery to its Revolutionary Impact on Live Cell Imaging

The jellyfish *Aequorea Victoria* is a bioluminescent inhabitant of the Pacific Ocean, emitting green light. In the early 1960s, Osamu Shimomura and colleagues found that the active component of *Aequorea* bioluminescence is the protein aequorin, emitting blue light in a Ca^{2+} dependent manner (Shimomura et al. 1962). The green appearance of *Aequorea* is due to a second protein, absorbing the blue light via radiationless Förster resonance energy transfer (FRET) and emitting it again in the green spectral range. This protein was consequently named green fluorescence protein (GFP) (Shimomura et al. 1962; Cormier et al. 1973). By proteolytic digestion of GFP, Shimomura could finally identify the actual chromophore as being a p-hydroxybenzylideneimidazolinone moiety (Shimomura 1979). The crucial breakthrough for the application of GFP in cell biology came along with the cloning of the gene by Douglas Prasher, more than 30 years after its initial discovery (Prasher et al. 1992). Martin Chalfie was the first to express GFP in other organisms than jellyfish, namely *E. coli* and *C. elegans* (Chalfie et al. 1994). Importantly, recombinantly expressed GFP retained its fluorescence properties, which implicated that the fluorescence does not require any other component specific to *Aequorea victoria*. Furthermore, GFP could be fused to proteins without influencing their activity, which demonstrated its virtually unlimited potential as a universal genetic tag in biological research (Wang and Hazelrigg 1994). Roger Tsien's laboratory demonstrated that chromophore formation by the side chains of Ser65-Tyr66-Gly67 is dependent on oxygen, since anaerobically expressed GFP did not fluoresce (Heim et al. 1994). Importantly, the fluorescence of such non-fluorescent GFP could be restored by adding oxygen. Based on these results, they proposed a chemical mechanism for chromophore formation with molecular oxygen as the only auxiliary factor. The same group engineered a series of point mutation variants with profound effects on the GFP spectral properties (Heim and Tsien 1996). A milestone on the way to rational design of superior GFP variants was the solution of the GFP crystal structure (Ormo et al. 1996; Yang et al. 1996). The nowadays most widely used enhanced GFP (eGFP) carrying several structure-guided mutations, is characterized by a higher brightness, less photobleaching and a single and amplified excitation peak. Recently, several other fluorescent proteins have been developed, most of them originating from the red fluorescent protein of *Discosoma*, covering a broad range of the visible spectrum (Shaner et al. 2005). The thus obtained variety of chromophores laid the cornerstone for the more complex bio-imaging techniques such as fluorescence resonance energy transfer (FRET) or induced photoactivation (Patterson and

Lippincott-Schwartz 2002; Zhang et al. 2002) used today. For their enormous contributions “to the discovery and development of the green fluorescent protein, GFP”, Osamu Shimomura, Martin Chalfie and Roger Tsien got awarded with the Noble Prize in Chemistry in 2008.

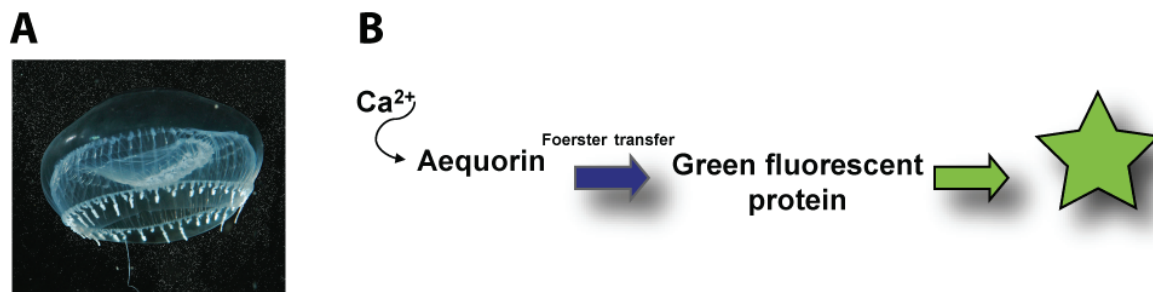


Figure 25: Bioluminescence of the jellyfish *Aequorea Victoria*.

(A) The jellyfish *Aequorea Victoria* in its natural habitat, the Pacific Ocean (Wikimedia Commons) and (B) scheme of the origin of its green fluorescence.

6.2 Green Fluorescent Protein: Biophysical and Structural Properties

As was already deduced from comparing the spectral properties of a fluorescent GFP peptide fragments to spectrometric properties of reference substances, the GFP chromophore contains a p-hydroxybenzylidenimidazolinone moiety (Shimomura 1979). It is formed out of the amino acid residues Ser65-Tyr66-Gly67 in the wild-type protein (Prasher et al. 1992). **Figure 26** shows the suggested mechanism of chromophore formation (Heim et al. 1994).

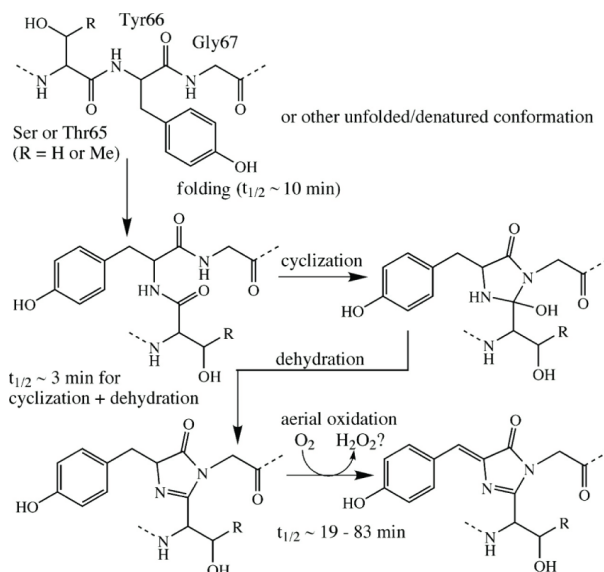


Figure 26: Catalytic mechanism of GFP chromophore formation.

Schematic illustration of the intramolecular chromophore biosynthesis with the estimated rate constants (Tsien 1998).

After having folded into an almost native conformation, the imidazolinone is formed by nucleophilic attack of the amide of Gly67 on the carbonyl of residue 65, followed by rehydration. In a final step, molecular oxygen dehydrogenates the α - β bond of residue 66 to put its aromatic group into conjugation with the imidazolinone. Only at this stage does the chromophore acquire visible absorbance and fluorescence. GFP has a quite complex absorption and emission spectrum (Heim et al. 1994). The original wild type GFP is characterized by a dual-peak excitation spectrum with a major absorption maximum at 395 nm and a minor one at 477 nm (**Figure 27**). Excitation at either wavelength results in the emission of green fluorescence at around 507 nm. The dual absorption of GFP is due to two alternative states of the chromophore, which are interconvertible by photoisomerization. The neutral phenol state absorbs at 395 nm whereas the deprotonated phenolate anion absorbs at 477 nm (Heim et al. 1995). Because of light-induced anionization, the neutral chromophore gives emission in a similar range as the anionic chromophore as has been shown by picosecond spectroscopy (Chattoraj et al. 1996). Mutations within the chromophore environment have a great impact on the spectral properties of GFP. The mutation Ser65T, which among other mutations is incorporated in the eGFP variant, leads to a suppression of the 395 nm excitation peak while the peak at longer wavelength is increased five- to sixfold in amplitude and shifted to 490 nm (**Figure 27**) (Heim et al. 1995).

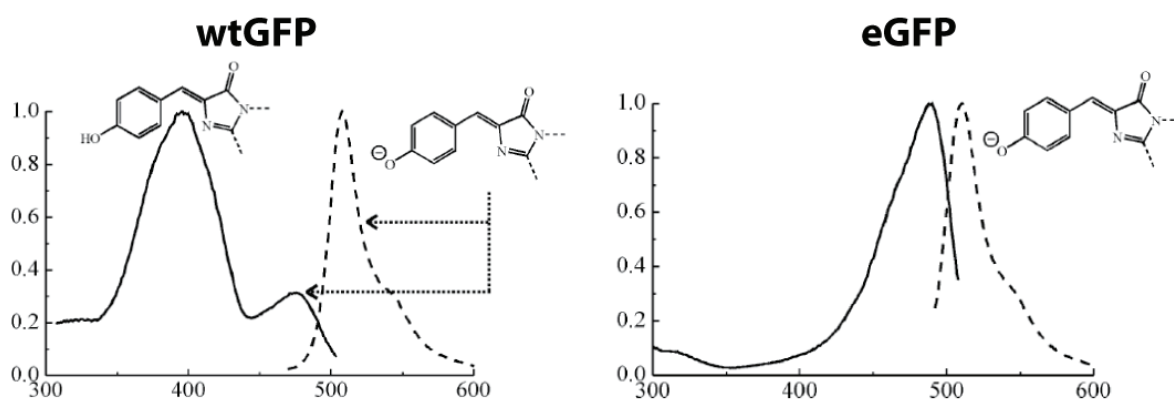


Figure 27: GFP excitation and emission spectra.

Excitation spectra of wild-type GFP (wtGFP) and enhanced GFP (eGFP) are shown as solid- and emission spectra as dotted line. The respective state of the chromophore is depicted above (Tsien 1998).

Although GFP had already been crystallized in the 1970s (Morise et al. 1974), the crystal structure of GFP was only reported 26 years later (Ormo et al. 1996; Yang et al. 1996). GFP consist of 238 amino acids, but only 229 residues are visible in the structure. Residues 1 and 230-238 were too disordered to be resolved. The protein folds into an 11-stranded β barrel which is wrapped around a single central helix. The barrel forms a cylinder with a length of 42 Å and a diameter of 24 Å (**Figure 28A**). The ends of the cylinder are capped by short, distorted helical segments and loops. The chromophore is integrated in the central helix,

thereby being completely protected from bulk solvent. The p-hydroxybenzylideneimidazolinone is embedded into a complex network of hydrogen bonds and polar interactions with surrounding amino acids and water molecules (**Figure 28B and C**). Comparison of the wtGFP structure to the S65T mutation gave a structural rationale for the observed changes in the excitation spectrum (Brejc et al. 1997). The main differences between both structures are found in the chromophore environment. The observed differences are due to the presence of an additional methyl group in the S65T mutation in comparison to the wild-type. The authors suggest a complicated rearrangement of the hydrogen bond network (Brejc et al. 1997). As a consequence, the negative charge of the phenolate state of the chromophore is stabilized in the S65T mutant, leading to the described spectroscopic changes.

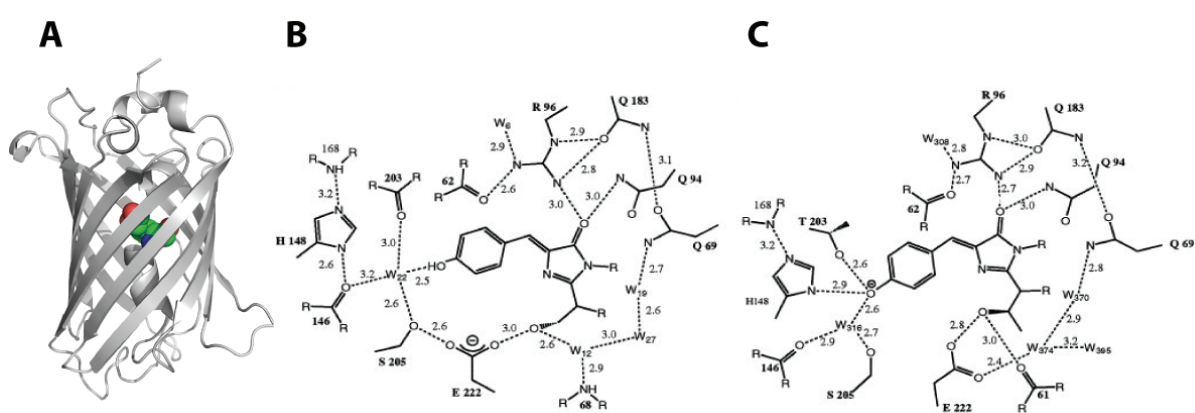


Figure 28: Structure of GFP and its integrated chromophore.

(A) Ribbon presentation (grey) of the GFP β -barrel (Yang et al. 1996) with the chromophore shown as colored spheres. Structural comparison of the chromophore environment of the wtGFP (B) and comparison to the GFP S65T mutation (C) according to Brejc. et al. 1997.

6.3 Nanobodies as a Versatile Tool for Specific Protein Targeting

The field of recombinant antibody technology has rapidly progressed over the last decade mainly driven by the ability of antibodies to specifically bind therapeutically as well as diagnostically relevant targets. However, conventional antibodies (immunoglobulins G - IgGs) have a major drawback, in that their recombinant expression is complicated by them being composed of four polypeptide chains (two heavy and two light chains, **Figure 29**), and being subject to extensive post-translational modification *in vivo*. Furthermore, their application within cells is limited by the impaired disulfide formation in the reducing environment of the cytosol, difficult subcellular targeting and inefficient assembly of epitope-recognizing parts of the variable light and heavy chain (Biocca et al. 1991; Cattaneo and Biocca 1999). A milestone was the discovery that antibody fragments, which consist of the N-terminal variable domains of heavy and light chains only (**Figure 29**), are sufficient for antigen binding

(Sundberg and Mariuzza 2002). However such Fab fragments remain difficult to express in microbial systems. Therefore, the discovery that camelids (bactrain camels, dromedries and llamas) produce functional antibodies devoid of light chains was a clear breakthrough in the field (Hamers-Casterman et al. 1993). Heavy-chain antibodies (HCAs) recognize their antigens via a single variable domain (referred to as VHH nanobody, **Figure 29**), which presents the smallest intact antigen binding fragment (~ 15 kDa) occurring in nature (Muyldermans 2001). The compactness of nanobodies has multiple advantages over conventional IgGs and smaller Fab fragments: (1) Only one domain has to be cloned and expressed, (2) specific nanobodies can easily be selected by phage display libraries and (3) nanobodies are highly soluble and stable and can be efficiently expressed in heterologous systems such as *E. coli* (Muyldermans et al. 1994). The combination of these properties make nanobodies highly interesting molecules for diagnostic as well as therapeutic purposes. Described diagnostic applications in cell biology include targeting and tracing of antigens in live cells, targeted modulation of enzymes as well as the usage as immobilized nanotraps to precipitate protein complexes *in vivo* and *in vitro* (Muyldermans 2001; Revets et al. 2005; Rothbauer et al. 2006; Rothbauer et al. 2008). Nanobodies are also discussed as novel agents for cancer therapy and other therapeutic applications (Cortez-Retamozo et al. 2004; Revets et al. 2005). Very recently, the first VHH successfully completed phase I clinical trials (www.ablynx.com).

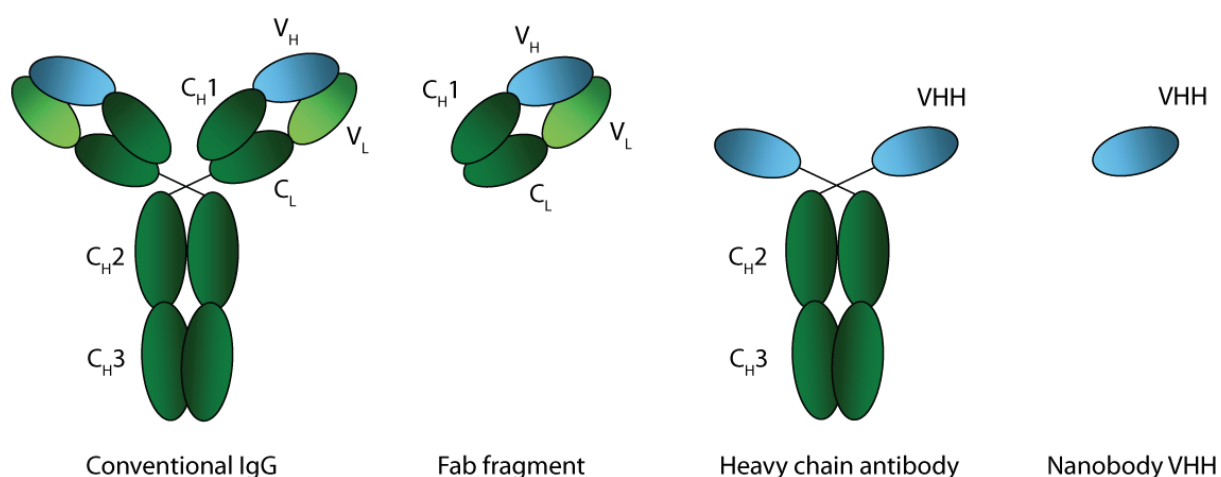


Figure 29: Schematic illustration of antibodies and fragments thereof.

Conventional IgGs are composed of two heavy and two light chains. The heavy chain harbors three conserved domains (C_H) and one variable domain (V_H). The light chain is composed of one variable (V_L) and one conserved domain (C_L). Fab fragment are monovalent antigen recognition domains resulting from proteolytic cleavage of conventional IgGs. Heavy chain antibodies are composed of two heavy chains only. The monovalent antigen recognition motif is formed by the VHH domain, also referred to as nanobody.

6.4 Structural Properties of Nanobodies

Determination of crystal structures of VHH nanobodies alone and in complex with antigen have elucidated several interesting structural features (Desmyter et al. 1996; Spinelli et al. 1996; Desmyter et al. 2002). Similar to conventional VH domains, nanobodies contain four framework regions (FRs) that fold into the typical immunoglobulin fold, and three complementary-determining regions (CDRs) that are involved in antigen binding (**Figure 30**). The FRs share a high sequence conservation with conventional antibodies and their 3D structures are superimposable (Muyldermans et al. 2001). However, four amino acids (Val37, Gly44, Leu45, Trp47) that are extremely conserved in VHs are constitutively substituted in the nanobodies (**Figure 30**). In conventional antibodies these residues form the hydrophobic core between the VH and the VL domain (VL site). In HCABs, they are mutated to more hydrophilic residues due to their exposed position. Further differences between VHs and nanobodies are reasoned by the fact, that in the absence of the VH-VL combinatorial diversity, new mechanism must account for a similar variability. For HCABs this is achieved in two ways; first the VL site provides an additional surface, which can be employed for antigen recognition. Furthermore, in comparison to VHs of human or mice, the variability of the CDR1 and CDR2 is increased. The CDR3 of a VHH is longer than that of VHs and also more accessible for solvent, thus creating a larger surface for antigen interaction. On the basis of these structural findings, several advantages of nanobodies can be rationally explained. The increased solubility, stability and efficient expression are most likely due the decrease in hydrophobicity at the VL site. Furthermore, nanobodies have been described to recognize recessed antigenic sites such as enzyme active centers, which can normally not be contacted by conventional antibodies (Lauwereys et al. 1998; De Genst et al. 2006). This feature has been attributed to their smaller size and the ability of the extended CDR3 loop to penetrate into such sites. Importantly, it facilitates their use as highly specific enzyme inhibitors.

A

	110	120	30	abc	140	50	abcdef	60	
RR6-R2	QVQLQESGGGLVQAGGSLR	LSCAASGR	RATSGHGHYGMGW	FRQVPGK	EREFVA	AI	RW-----	SGKETWYKDS	
RR6-R9	QVQLQESGGGLVQAGESL	KLSCAASG	NTFSG---	GFMGW	FRQAPGK	QREL	VATIN-----	SRGITNYADF	
hCG-H14	QVQLQESGGGLVQAGGSL	RLSCAASG	RTGST---	YDMGW	FRQAPGK	ERESV	AAINW-----	DSARTYYASS	
cAb-LYS3	DVQLQASGGGSVQAGGSL	RLSCAASG	YTIGP---	YCMGW	FRQAPGK	EREGV	AAINM-----	GGGITYYADS	
cAb-RN05	QVQLVESGGGLVQAGGSL	RLSCAASG	YAYTY---	IYMGW	FRQAPGK	EREGV	AAMDS-----	GGGGTLYADS	
cAb-CA05	QVQLVESGGGSVQAGGSL	RLSCAASG	YTVST---	YCMGW	FRQAPGK	EREGV	ATIL-----	GGSTYYGDS	
AMYL 07	QVQLVESGGGSVQAGGSL	RLSCAASG	YTFSS---	YPMGW	FRQAPGK	EC	ELVSRIF-----	SDGSANYAGS	
AMYL D10	DVQLVESGGGTVPAGGSL	RLSCAASG	NTLCT---	YDMSW	YRRAPGK	GRDF	VSGID-----	NDGTTYVDS	
Pot VH	EVHLLSEGGN	LVQPPGSL	RLSCAASG	FTFNI---	FVMSW	VRQAPGK	GLEWVS	GVFG-----	SGGNTDYADA

	70	80	abc	90	100	abcdefghijklmno	p	110	
RR6-R2	VKGRFTISRDN	AKTTVYLQ	MNSLKP	PEDTAVYY	CAAR	PVRVDD	ISL	PVGF-----	DYWGQGTQVTVSS
RR6-R9	VKGRFTISRDN	AKTTVYLE	MNSLE	PEDTAVYY	CYTHY	FR-----			SYWGQGTQVTVSS
hCG-H14	VRGRFTISRDN	AKTTVYLQ	MNSLKP	PEDTAVYT	CGAGE	GGTW-----			DSWGQGTQVTVSS
cAb-Lys3	VKGRFTISRDN	AKNTVYLL	MNSLE	PEDTAIYY	CAAD	STIYAS	YEC	GHGLSTGGYGYDS	SWGQGTQVTVSS
cAb-RN05	VKGRFTISRDK	GKNTVYLQ	MDSLKP	PEDTATYY	CAAGG	YELRDR	TY-----		GQWGQGTQVTVSS
cAb-CA05	VKGRFTISRDN	AKNTVYLQ	MNSLKP	PEDTAIYY	CAGS	TVA	STGWC	SRLRPYDY-----	HYRGQGTQVTVSS
AMYL 07	VKGRFTISRDN	AKNTAYLQ	MDSLKP	PEDTAVYY	CAAGP	SGKLVV	AGRT	CYGP-----	NYWGQGTQVTVSS
AMYL D10	VKGRFTISQGN	AKNTAYLQ	MDSLKP	DDTAMYY	CKPS	LRYGL	PGC	PI-----	IPWGQGTQVTVSS
Pot VH	VKGRFTITRD	NSKNTLYLQ	MNSLRA	EDTAIYY	CAK	HRVSYV	LTGF-----		DSWGQGTQVTVSS

B**Figure 30:** Structure-based alignment of nanobodies.

(A) Alignment of nanobody sequences with known crystallographic structure with a human VH domain (Pot VH). The CDRs are shown in red (CDR1) green (CDR2) and purple (CDR3). Cysteines which are involved in disulfide bridge formation are highlighted in yellow and framework regions are colored grey. Adopted from Muyldermans 2001. VHH specific residues in the VL site are shown in magenta. (B) Crystal structure of the amylose D10 nanobody (AMYL D10) (Desmyter et al. 2002). Structural motifs are shown in similar color code as described for (A).

6.5 Objectives

While both improving GFP-fluorescence and developing spectral derivatives thereof have led to a dramatic upsurge in the use of GFP in cell biology, the possibility to inducibly modulate the fluorescence of GFP *in vitro* and *in vivo* is still lacking. To overcome these drawbacks, the influence of several GFP specific nanobodies on the GFP spectral properties was tested. Finally, two GFP-specific nanobodies with opposing effects on GFP fluorescence were devised: one suppresses, the other enhances the fluorescence.

The goal of this thesis was to provide a structural understanding of the observed modulation of GFP fluorescence. To this end, high resolution structures of the two GFP nanobody complexes should be solved by X-ray crystallography. A detailed spectroscopic analysis should augment the conclusion eventually derived from the structural analysis. Furthermore, in collaboration with the group of Prof. Dr. Heinrich Leonhardt and Dr. Ulrich Rothbauer, the practical benefit and potential applications of GFP fluorescence modulating nanobodies in living cells should be demonstrated.

7 Materials and Methods

The following chapter lists experimental details of the production of GFP-nanobody complexes and their structural and spectroscopic characterization. General materials and methods are not mentioned here but can be found in Part I of this thesis.

7.1 Materials

7.1.1 Chemicals

Crystallization screens, crystallization grade reagents and crystallization tools were obtained from Hampton Research (Aliso Viejo, USA), Qiagen, (Hilden, Germany) and Jena Bioscience (Jena, Germany). Chemicals for crystallization were of the highest possible grade and were purchased from Merck (Darmstadt, Germany) or Sigma (Deisenhofen, Germany).

7.1.2 Bacterial Strains

Table 11: Bacterial Strains

<i>E.coli</i> strain	Genotype	Source
B834 Rosetta (DE3)	F ⁻ <i>ompT hsdSB(rB⁻ mB⁻) met gal dcm</i> (DE3)	Novagen, Madison USA

7.1.3 Preparation of Minimal Medium for Selenomethionine Expression

All amino acids were reagent-grade L-enantiomers purchased from Sigma (Deisenhofen, Germany). After dissolving the components (**Table 12**) in 2000 ml FPLC grade water, solution A was autoclaved, cooled to 37°C and supplemented with 200 ml filter-sterilized solution B plus 125 mg selenomethionine (Calbiochem, Schwalbach, Germany).

Table 12: Composition of LeMaster's medium (Hendrickson et al. 1990)

Autoclavable solution A (2000 ml)	
Alanine 1.0 g	Serine 4.166 g
Arginine hydrochloride 1.16 g	Threonine 0.46 g
Aspartic acid 0.8 g	Tyrosine 0.34 g
Cystine 0.066 g	Valine 0.46 g
Glutamic acid 1.5 g	Adenine 1.0 g
Glutamine 0.666 g	Guanosine 1.34 g
Glycine 1.08 g	Thymine 0.34 g
Histidine 0.12 g	Uracil 1.0 g
Isoleucine 0.46 g	Sodium acetate 3.0 g
Leucine 0.46 g	Succinic acid 3.0 g
Lysine hydrochloride 0.84	Ammonium chloride 1.5 g
Phenylalanine 0.266 g	Sodium hydroxide 1.7 g
Proline 0.2 g	Dibasic potassium phosphate 21.0 g
Non-autoclavable solution B (200 ml)	
Glucose 20.0 g	Conc. sulfuric acid 16.0 μ l
Magnesium sulphate 0.5 g	Thiamine 10.0 mg
Iron sulphate 8.4 g	

7.2 Methods

7.2.1 Bioinformatic Methods

7.2.1.1 Structure Visualization and Analysis

Calculation of buried surface areas was performed with the protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (Krissinel and Henrick 2007). Superpositions of structures and calculation of root mean square deviation RMSD values were conducted using the CaspR RMSDcalc web-server (Claude et al. 2004). Images of crystal structures were prepared with PyMol (<http://www.pymol.org>).

7.2.2 Protein Biochemistry Methods

7.2.2.1 Protein Expression

Nanobodies were expressed in Rosetta cells transformed with expression vector pHEN6 containing the coding sequence of the GFP nanobody as described previously (Rothbauer et

al. 2008). Native GFP and eGFP cloned into the pRset5D expression plasmid were expressed accordingly (Rothbauer et al. 2008). Selenomethionine-containing GFP was expressed in *E. coli* strain B834 (Rosetta (DE3)) using LeMaster's medium supplemented with selenomethionine (**Table 12**).

7.2.2.2 Protein Purification

All buffers and their components used for protein purification are listed in **Table 7**.

For purification of C-terminally His₆-tagged *Enhancer* or *Minimizer*, bacterial cell material was resuspended in 10 ml of lysis buffer and then sonified. After centrifugation soluble extracts were loaded on a pre-equilibrated 1 ml HiTrap-column (GE Healthcare; Freiburg, Germany). The nanobodies were eluted by a linear gradient ranging from 20 mM (Ni-NTA washing buffer) to 500 mM imidazol (Ni-NTA elution buffer). Elution fractions containing the *Enhancer* and *Minimizer*, respectively, were pooled and further purified by gel filtration on a Superdex 75 column (GE Healthcare; Freiburg, Germany) in gelfiltration buffer. For complex purification of wtGFP-Enhancer and wtGFP-Minimizer complexes, purified nanobodies were pre-bound via their C-terminal His₆-tag to a 1ml HiTrap-column (GE-Healthcare Life Sciences). Subsequently, wtGFP was isolated from crude cell extracts by binding to the pre-charged column. Complexes were eluted by increasing concentration of imidazol and separated from unbound protein by gel filtration chromatography using a Superdex 200 column (GE Healthcare; Freiburg, Germany) in gelfiltration buffer.

Table 13: Protein purification buffers

Buffer	Composition
Lysis buffer	500 mM NaCl, 12 mM Phosphate, 2.7 mM KCl, pH 7.4, 20 mM Imidazol, 1 mM PMSF
Ni-NTA washing buffer	500 mM NaCl, 12 mM Phosphate, 2.7 mM KCl, pH 7.4, 20 mM Imidazol
Ni-NTA elution buffer	500 mM NaCl, 12 mM Phosphate, 2.7 mM KCl, pH 7.4, 500 mM Imidazol
Gefiltration buffer	20 mM Tris 7.5, 100 mM NaCl

7.2.3 Fluorescence Spectroscopy

Fluorescence assays were performed either by scanning a 96 well microplate (Nunc) on a Typhoon Trio (GE Healthcare, Freiburg, Germany; excitation: 488 nm, emission filter settings: 520 nm BP 40) or using a monochromator-based microplate reader (Infinite M1000, Tecan, Mainz, Germany; excitation wtGFP at 395 nm or 475 nm, excitation eGFP: 488 nm and emission 507 nm +/- 10 nm). Increasing amounts of the respective GFP binding nanobody were titrated to 100 µl of 50 nM GFP in gelfiltration buffer.

GFP fluorescence excitation spectra were recorded with a FluoroMax-P fluorimeter (HORIBA Jobin Yvon). Typically, 0.5 μM protein in gel filtration buffer was measured in a 1 ml quartz cuvette. The band pass was 5 nm for excitation and 5 nm for emission. Samples were excited at 395 nm and 475 nm, respectively, and excitation spectra were recorded in the range of 480 nm to 600 nm.

Fluorescence absorption spectra were recorded on a UV / Vis Spectrophotometer (Beckman Coulter). Absorption of 0.5 μM GFP alone and 0.5 μM GFP-*Enhancer* / -*Minimizer* complexes, respectively, was detected during continuous excitation steps (1 nm) from 250 nm – 700 nm.

All fluorescent data were analyzed using Microsoft Excel.

7.2.4 Structure Determination by X-Ray Crystallography

The following chapter should give a short introduction into X-ray crystallography with a focus on techniques employed during this thesis. A detailed description of the underlying theory would go far beyond the scope of this text and can be found elsewhere (Rhodes 2000; Drenth 2002; Giacovazzo 2003).

7.2.4.1 Theoretical Background of X-ray Diffraction

X-rays are electromagnetic waves, having a wavelength in the range of atomic distances (10^{-10} m = 1 Å), that interact with electrons in the outer sheets of atoms. The excited atom emits X-rays of the same wavelength in every direction by means of elastic scattering. Normally, the emitted waves scattered from different atoms interfere and cancel each other out. However, constructive interference can occur if the light path of the two waves differs by $n \cdot \lambda$ (n =integer) and a reflection can be recorded. Crystals are highly ordered three dimensional arrangements composed of unit cells related by translational symmetry. Therefore, X-rays diffracted from crystals can be described as a reflection from imaginary lattice planes. Intersections of the lattice planes with the unit cell axes are termed Miller Indices (h, k, l) and describe the orientation and the spacing between a set of parallel lattice planes. The conditions for constructive interference are given by Bragg's Law:

$$n \cdot \lambda = 2 \cdot d \cdot \sin \theta$$

Only if the distance d between the parallel lattice planes and the angle θ between the lattice planes (h, k, l) follow Bragg's law, a corresponding reflection (defined by the indices h, k, l) can be observed. With the knowledge of the positions of the reflections, the geometry of the unit cell can be determined.

7.2.4.2 The Phase Problem in Crystallography

To reconstitute a three-dimensional electron density distribution of a crystallized molecule from the information contained in the reciprocal diffraction pattern, several parameters of each reflected wave have to be determined. Mathematically, the three dimensional electron density distribution of a molecule ($\rho(x, y, z)$) can be described by the following equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} |F(h, k, l)| \cdot \exp[-2\pi i \cdot (hx + ky + lz - \alpha(h, k, l))]$$

This formula represents the summation over all structure factors $F(h, k, l)$ for each position (x, y, z) in a normalized unit cell ($1/V$). Several parameters for each diffracted wave are needed to calculate back to the three-dimensional model in real space via Fourier Transformation. The wavelength of the scattered rays is given by the wavelength of the incident beam. Whereas the amplitude and the position of the reflection in reciprocal space ($|F(h, k, l)|$) can be directly determined from the X-ray diffraction pattern, the phase of the scattered reflections ($\alpha(h, k, l)$) is lost during the measurement. This lack of information is called the “phase problem” in crystallography. Several methods have been developed to approximate the missing phase information. An important tool to obtain phase angles in *de novo* structure determination is the Patterson function. It comprises a Fourier Transformation of the measured intensities and does not depend on phases.

$$P(u, v, w) = 1/V \sum_{hkl} |F_{hkl}|^2 \cos(2\pi(hu + kv + lw))$$

The Patterson unit cell is characterized by u, v, w coordinates and has an identical dimension as the real unit cell. However, the Patterson map is comprised of interatomic distance vectors and is usually too complicated to be directly interpreted for the location of individual atoms. Still, as the peak heights at the endpoints of the vectors depend on the electron number of the respective atom, relative positions of heavy atoms in a unit cell can be determined.

7.2.4.3 Single and Multiple Wavelength Anomalous Dispersion (SAD/MAD)

Near the absorption edge of an atom resonance effects can be observed, if the energy of the incident beam is close to the Eigenfrequency. In the resonance range some photons are emitted with the same energy but with a phase shift. This phenomenon is described as anomalous dispersion. Single or multiple wavelength anomalous dispersion (SAD or MAD) depends on the presence of one or more heavy atoms in the polypeptide chain that cause significant anomalous scattering. Atoms in proteins which are suitable for this purpose are

sulfur or heavier atoms, for example metal ions in metalloproteins. The most commonly used atom for phase determination via MAD, however, is selenium, since it is usually possible to replace the natural sulfur containing amino acid methionine by selenomethionine. With the increased availability of synchrotron sources, which provide tunable X-ray beams over a broad spectrum of wavelengths, this technique has nowadays developed into the most widely used method for experimental phasing of protein structures.

While in normal scattering, the structure factors have the same amplitude and opposite phases according to Friedel's Law, anomalous scattering causes violation of Friedel's Law.

$$\Delta|F|_{ano} = (|F_{PH}(h, k, l)| - |F_{PH}(-h, -k, -l)|) \cdot \frac{f'}{2f''}$$

where $\Delta|F|_{ano}$ is the difference between the structure factors of the related Bijvoet pair $|F_{PH}(h, k, l)|$ and $|F_{PH}(-h, -k, -l)|$ scaled up by the factor $f'/2f''$. With help of a Patterson map calculated with $\Delta|F|_{ano}^2$ the amplitude of structure factor ($|F_H|$) and the phases (α_H) of the heavy atoms can be determined. Ultimately, the protein phase angle can be estimated (α_P) by the following equation,

$$F_{PH} = |F_P|\alpha_P + |F_H|\alpha_H$$

where F_{PH} is the structure factors of the derivate crystal and $|F_P|$ is the amplitude of the protein structure factor.

As anomalous scattering depends on the wavelength, usually three wavelengths (peak, inflection and remote) are recorded in a MAD experiment. In comparison to a SAD experiment, where anomalous data is recorded at a single wavelength only, MAD allows more precise phase estimation.

7.2.4.4 Molecular Replacement

Another way of solving the phase problem in X-ray crystallography is the molecular replacement method (MR). This however requires the existence of a structurally related search model. The search model can consist of already solved parts of the unknown structure; alternatively a close structural homolog can be used. Because of the increasing number of known structures deposited in the protein data bank (PDB), the latter approach becomes more and more feasible. In order to obtain useful phase angles of the unknown structure, the search model has to be oriented in the electron density of the unknown structure. The replacement is a six-dimensional search problem, which can be broken down

into a three-dimensional translation and a three-dimensional rotation search. In a primary rotation search the intramolecular vectors, which were derived from the Patterson function and depend on the orientation of the molecule only, are correlated with the search model. The consecutive translational search uses intermolecular cross vectors, which depend on both the orientation and the position of the molecule. The derived coordinates of the search model in the unit cell then allow the calculation of a new structure factor amplitude $|F_{calc}|$ and an estimation of the respective phase angles α_{calc} . The calculated phase angles α_{calc} can then be combined to measured structure amplitudes $|F_{obs}|$ resulting in an estimate of the structure factor.

$$F(h) = |F_{obs}|(h) \cdot \exp(i\alpha_{calc}(h))$$

The use of $\alpha_{calc}(h)$ in calculating the electron density map results in a bias towards the model structure employed to generate the phase angles. Therefore, an electron density map is typically obtained from structure factors which are weighted towards $|F_{obs}|$ to minimize model bias, according to the following equation.

$$F(h) = (2|F_{obs}|(h) - |F_{calc}|(h)) \cdot \exp(i\alpha_{calc}(h))$$

7.2.4.5 Crystallization and Structure Determination of the GFP-Enhancer and the GFP-Minimizer Complex

Purified GFP-nanobody-complexes were crystallized by hanging drop vapor diffusion. 1 μ l of protein solution at 10 mg/ml concentration was mixed with 1 μ l of the reservoir solution (native GFP-*Enhancer*: 0.1 Citrate pH 4.0, 1.6 M Am_2SO_4), (Selenomethionine (SeMet)-GFP-*Enhancer*: 60% (v/v) MPD, 100 mM NaAc pH 4.6, 10 mM CaCl_2), (GFP-*Minimizer*: 100 mM Mes pH 6.5, 30% (v/v) PEG8000, 15% (v/v) Glycerol).

Crystals of native and SeMet-GFP-*Enhancer* grew after one week at 20°C and were flash frozen in liquid nitrogen. Native GFP-*Enhancer* crystals were recorded at the X06SA beamline (Swiss Light Source) to 2.9 Å. Multiple wavelength diffraction data at the K absorption edge of selenium were measured at the same beamline to 2.15 Å. (**Table 14**).

GFP-*Minimizer* crystals occurred overnight and were flash frozen in liquid nitrogen the next day. Data collection was performed at the beamline ID29 at the European Synchrotron Radiation Facility to a resolution of 1.6 Å (**Table 15**).

Diffraction data of the complexes were processed with XDS (Kabsch 1993). The SeMet-GFP-*Enhancer* and GFP-*Minimizer* crystals belong to space groups $P4_22$ and $P2_12_12_1$,

respectively, with one (*GFP-Minimizer*) or two (*SeMet-GFP-Enhancer*) complexes in the asymmetric unit. In the case of the *GFP-Enhancer* structure, 3 selenium sites per complex were located with autoSHARP (Global Phasing, Cambridge). Single wavelength anomalous dispersion phasing and solvent flipping yielded an interpretable experimental electron density map. The models for *GFP* and *Enhancer* were built with COOT (Emsley and Cowtan 2004) and refined with PHENIX (Adams et al. 2002), using overall isotropic B-factors and bulk solvent corrections, individual B-factor refinement, simulated annealing, crystallographic- and positional refinement. The *GFP-Minimizer* structure was determined by molecular replacement with Phaser using an individual *GFP* and a nanobody polypeptide chain from the previously determined *GFP-Enhancer* structure as independent search models. The replacement model was manually altered with COOT and refined using similar procedures as described for the *GFP-Enhancer* structure.

Coordinates and structure factors of the *GFP-Enhancer* and *GFP-Minimizer* complex were deposited to the Protein Data Bank (PDB) with accession number 3K1K and 3G9A, respectively.

8 Results

8.1 GFP-Binding Nanobodies Modulate GFP Fluorescence

The group of Prof. Dr. Leonhardt and Dr. Rothbauer and has great expertise in the generation of VHH nanobodies and their application in cell biology (Rothbauer et al. 2006; Rothbauer et al. 2008; Zolghadr et al. 2008). In foregoing work, seven VHH nanobodies from alpacas with strong binding to GFP were selected by phage display. The resulting GFP binding proteins (GBPs 1 – 7) were cloned with a C-terminal histidine (His₆)-tag, expressed in *Escherichia coli* and purified by immobilized metal affinity chromatography (IMAC).

Because it was anticipated that nanobody binding might alter GFP fluorescence properties, a fluorescent *in vitro* binding assay was performed by titrating increasing amounts of GBP1 - 7 to wild type GFP (wtGFP). Fluorescence intensities were recorded in a 96 well microplate format and subsequently quantified (**Figure 31A** and **B**). Strikingly, two nanobodies, GBP1 and GBP4, indeed had a pronounced effect on the fluorescence emission of wtGFP. Whereas binding of GBP1 leads to a 4-fold enhancement, binding of GBP4 drastically reduces the fluorescence by a factor of 5. Overall, there is a remarkable 20 fold difference in fluorescence between the two nanobody complexes under the conditions applied. Attributing to their observed impact on GFP fluorescence GBP1 and GBP4 were termed *Enhancer* and *Minimizer*, respectively.

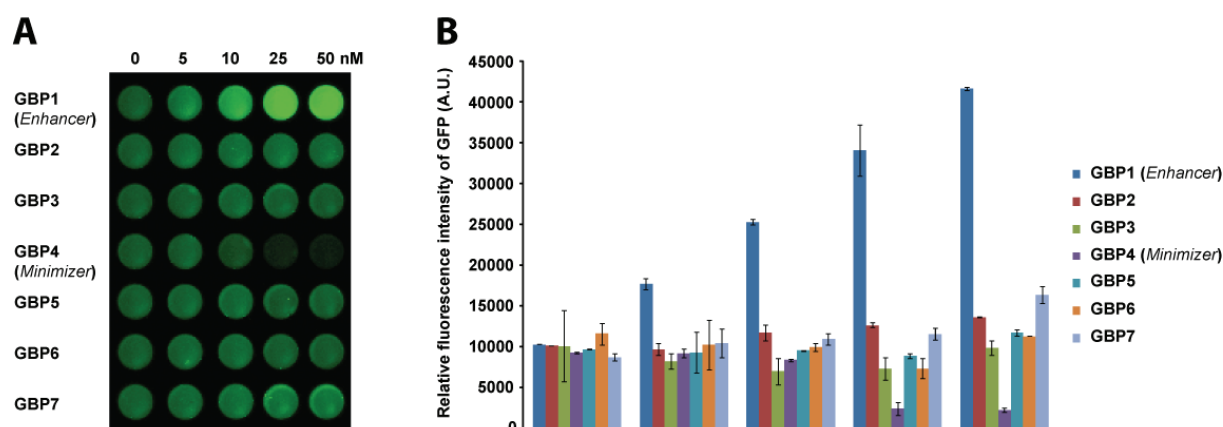


Figure 31: Identification of fluorescence modulating nanobodies.

(A) Titration of seven unique GFP binding nanobodies (GBP1 – 7, 0 – 50 nM) on purified GFP (50 nM / well). After complex formation, the signal intensity of GFP was detected by laser scanning (excitation 488 nm) (B) Quantification of the measurements shown in (A) with +/- standard deviation (S.D.) depicted as black bars.

The augmented fluorescence of the *Enhancer*–GFP complex is comparable to the improved spectral properties of eGFP. This led to the question, whether the *Enhancer* might be able to even further increase the optimized fluorescence of eGFP. Indeed, binding of *Enhancer* to eGFP resulted in an additional fluorescence increase of about 1.5–fold. In contrast, binding of *Minimizer* dramatically reduced the fluorescence of eGFP by a factor of 8 (**Figure 32**).

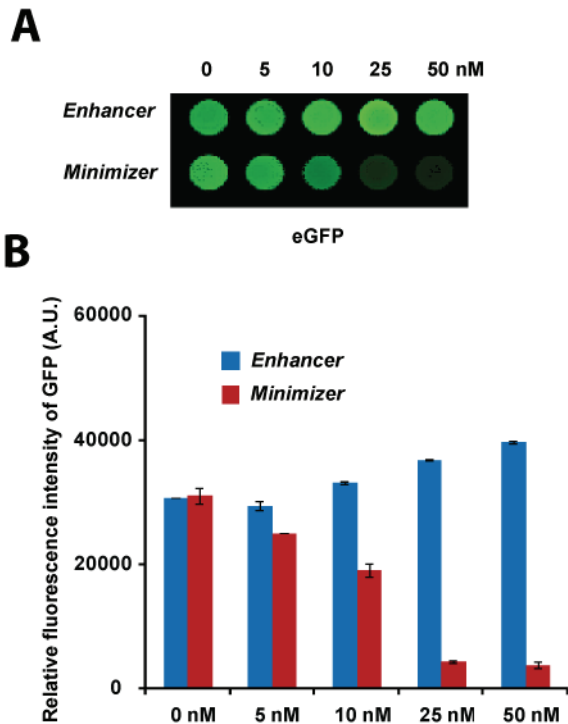


Figure 32: Effect of *Enhancer* and *Minimizer* on eGFP fluorescence.

(A) Titration of *Enhancer* and *Minimizer*, respectively (0 – 50 nM) on purified eGFP (50 nM / well). After complex formation, the signal intensity of GFP was detected by laser scanning (excitation 488 nm) (B) Quantification of the measurements shown in (A) with +/- S.D. depicted as black bars.

Based on the opposing effects of *Enhancer*- and the *Minimizer*-binding it was tested how GFP fluorescence is modulated in the presence of both nanobodies (**Figure 33A**). Interestingly, after a primary addition of *Enhancer*, GFP fluorescence increases, and is only slightly reduced by the consecutive addition of *Minimizer*. However, when *Minimizer* is added first, the reduction of fluorescence can be completely reversed by subsequent addition of *Enhancer* (**Figure 33B**).

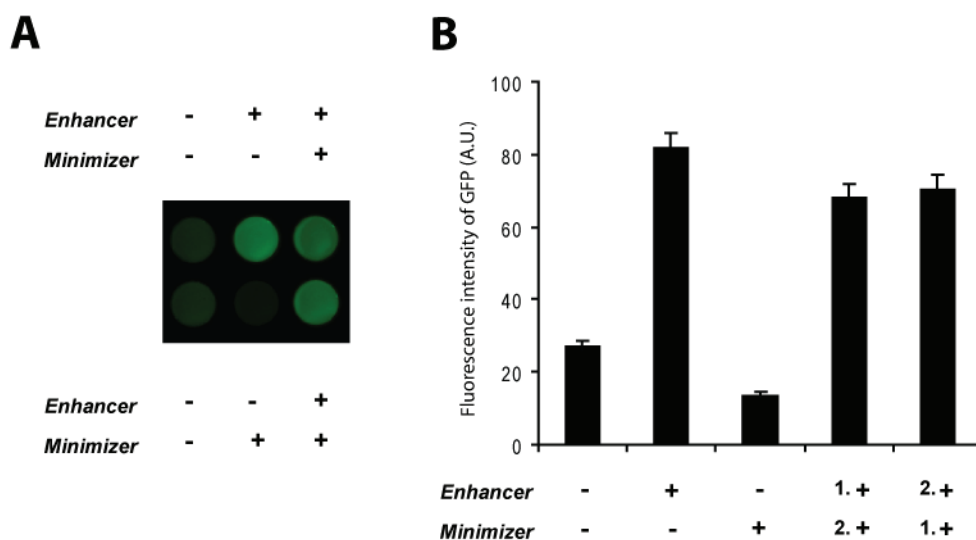


Figure 33: Modulating GFP fluorescence with nanobodies.

(A) *Upper row:* wtGFP (50 nM) was either mock incubated, incubated with equimolar amounts of *Enhancer* or *Enhancer* was added followed by the addition of equimolar amounts of *Minimizer*. *Lower row:* same experimental set up as above but with *Minimizer* being added first. (B) Quantification of fluorescence data. Represented here are means of 3 independent experiments +/- S.D. The order of the sequential incubation with *Enhancer* and *Minimizer* is indicated below.

Thus, binding of both nanobodies seems to be mutually exclusive. Although the ability of *Enhancer* to replace *Minimizer* at equimolar concentrations indicates a higher affinity, no significant difference in binding constants (K_d s) could be detected. Both nanobodies have a $K_d < 1 \mu\text{M}$ as was determined by Attana continuous flow measurements (data not shown).

8.2 Purification and Crystallization of the GFP-*Enhancer* and GFP-*Minimizer* Complexes

Prerequisite for protein crystallography is a highly pure protein sample. The GFP-*Enhancer* and GFP-*Minimizer* complex were constituted by first coupling the GFP binding nanobodies to Ni-NTA material via their C-terminal His₆-tags and then using the pre-charged material to capture GFP in a second affinity step. To separate the GFP-nanobody complexes from an excess of unbound nanobodies, a final gel filtration purification step was performed. Whereas the GFP-*Enhancer* preparation contained an excess of unbound nanobodies which eluted at a higher retention volume the GFP-*Minimizer* preparation eluted in a single peak corresponding to a 1:1 complex (**Figure 34**). In both cases, the complex fractions (**Figure 34**; peak 1) were pooled and concentrated to ~10 mg/ml.

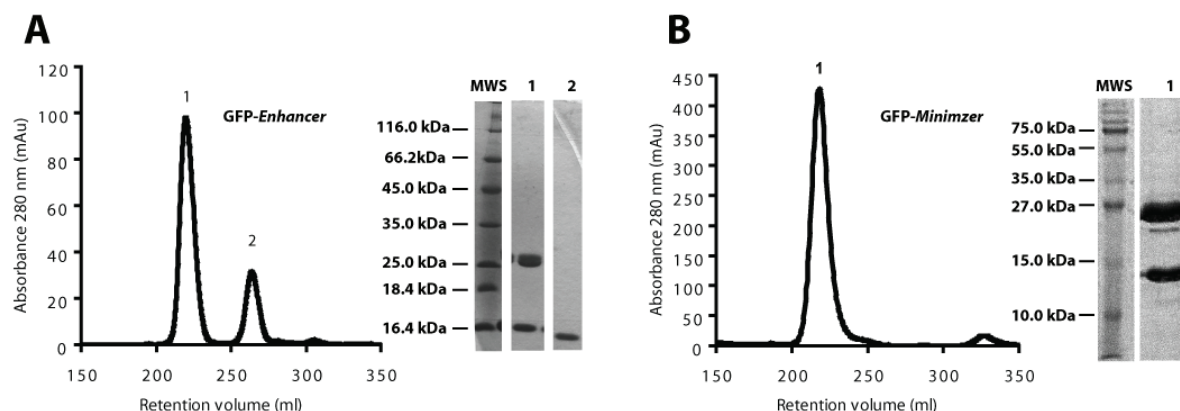


Figure 34: Purification of GFP nanobody complexes.

Gelfiltration profiles and SDS-PAGE analysis using Coomassie staining of (A) the GFP-*Enhancer* and (B) the GFP-*Minimizer* complex, respectively. (A) Excess of *Enhancer* after Ni-NTA affinity purification could be separated from GFP-*Enhancer* complex (peak 2 and 1, respectively). Molecular weight standards (MWS) for SDS-PAGE are indicated. (B) GFP-*Minimizer* complex preparation is stoichiometric (1:1) after Ni-NTA affinity purification and results in a single peak.

The concentrated protein samples were used for initial crystallization screening. Small needles of the GFP-*Enhancer* were found in one particular crystallization condition (**Figure 35A**) after several days of incubation. By evaluating the effects of a series of additives, the original condition could be significantly improved. The crystals now grew to substantially bigger size (100 x 200 x 10 μm) (**Figure 35B**). Nevertheless, these crystals were quite flat in shape and stayed mostly attached to the surface of the well which hindered crystal transfer. The best crystals were obtained when using complexed GFP without C-terminal His₆-tag. Because of the nanobody mediated capturing step during complex purification, the His₆-tag on GFP became dispensable. The improved GFP-*Enhancer* complex reproducibly crystallized in freely floating hexagons of ~ 100 x 100 x 70 μm in dimension (**Figure 35C**). Before flash freezing the crystals in liquid nitrogen, the crystallization buffer was supplemented with 20% (v/v) glycerol. As became obvious during subsequent structure determination, the resulting diffraction data was insufficient for a complete structural refinement (**chapter 8.4**). Therefore *Enhancer*-complex containing selenomethionine labeled GFP was employed in further crystallization trials. Surprisingly, substitution of methionine to selenomethionine within GFP dramatically changed the conditions in which GFP-*Enhancer* crystals could be obtained. Whereas previous crystals grew in high concentrations of ammonium sulfate, the selenomethionine containing complex crystallized in 60% (v/v) 2,4-methylpentadiol (**Figure 35D**). Notably, these crystals were also different in shape, growing as cubes of a size of ~ 50 x 50 x 50 μm . The crystals were directly flash frozen and stored in liquid nitrogen until measurement.

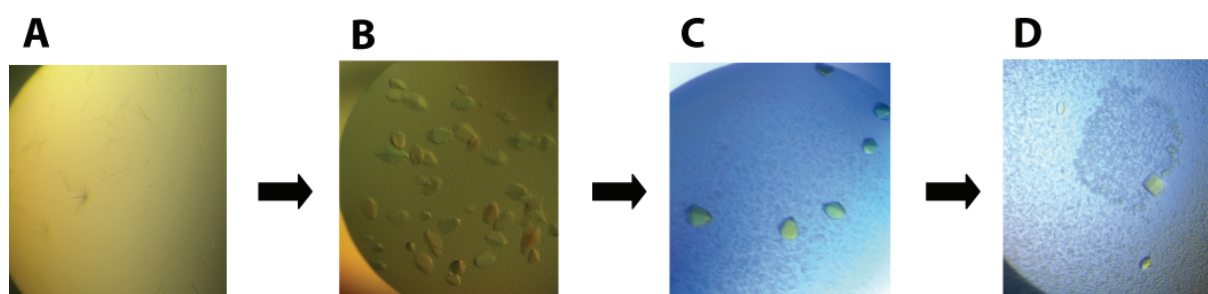


Figure 35: Crystal refinement of the GFP-*Enhancer* complex.

(A) Initial small needles obtained in commercial screens (0.1 M NaAc 5.0, 2M Am₂SO₄). (B) Improved crystals after additive screen of the initial condition (0.1 M NaAc 5.5, 1.7 M Am₂SO₄, 0.1 M Am₂HCitrate). Crystals grew as two-dimensional thin plates. (C) Further improved hexagonally shaped crystals resulting from new GFP construct design (buffer condition: 0.1 Citrate 4.0, 1.6 M Am₂SO₄). (D) GFP-*Enhancer* crystals of selenomethionine labeled GFP grew in a new, chemically deviant condition (60% (v/v) MPD, 100 mM NaAc 4.6, 10 mM CaCl₂).

In contrast to the quite complicated crystallization of the GFP-*Enhancer* complex, crystals of the GFP-*Minimizer* occurred overnight in several conditions of the initial screen. **Figure 36** shows an image of crystals which were subsequently employed for data collection. The crystals grew in branched bars of 50 x 20 x 20 μm in dimension. Strikingly, the fluorescence modulation effect of nanobody binding to GFP fluorescence which was first observed in solution was also visible in the crystal state. Interestingly and reflecting their fluorescence properties, GFP-*Enhancer* crystals appeared in a bright green, while GFP-*Minimizer* crystals were almost colorless (**Figure 35**, **Figure 36**)

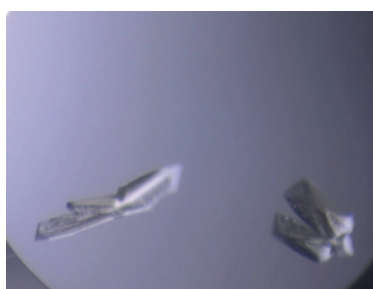


Figure 36: Crystals of the GFP-*Minimizer* complex.

Crystals obtained in commercial screens grew already in cryo conditions (100 mM Mes 6.5, 30% (v/v) PEG8000, 15% (v/v) Glycerol) and were subjected to data collection after flash freezing.

8.3 Data Collection

Native as well as selenomethionine-containing crystals of the GFP-*Enhancer* complex were measured at the X06SA beamline (Swiss Light Source) using a PILATUS 6M detector. Native data sets were recorded over an oscillation range of 90° (oscillation rate 1°) to a resolution of 2.9 Å and at a wavelength of 1.00 Å. Selenomethionine-containing crystals were measured at the selenium K edge. The optimal wavelengths were determined with a fluorescence scan prior to data collection (**Figure 37A**). Finally, data at the peak wavelength (12.66007 keV, $f' =$

-8.350 and $f'' = 6.102$), the inflection point (12.65760 keV, $f' = -10.857$ and $f'' = 3.648$) and the high remote (12.758206 keV) were collected.

Native data of GFP-*Minimizer* crystals were collected at the beamline ID29 at the European Synchrotron Radiation Facility over an oscillation range of 90° (oscillation rate 1°) to a resolution of 1.6 \AA and at a wavelength of 0.98137 \AA using an ADSC QUANTUM Q315R area detector.

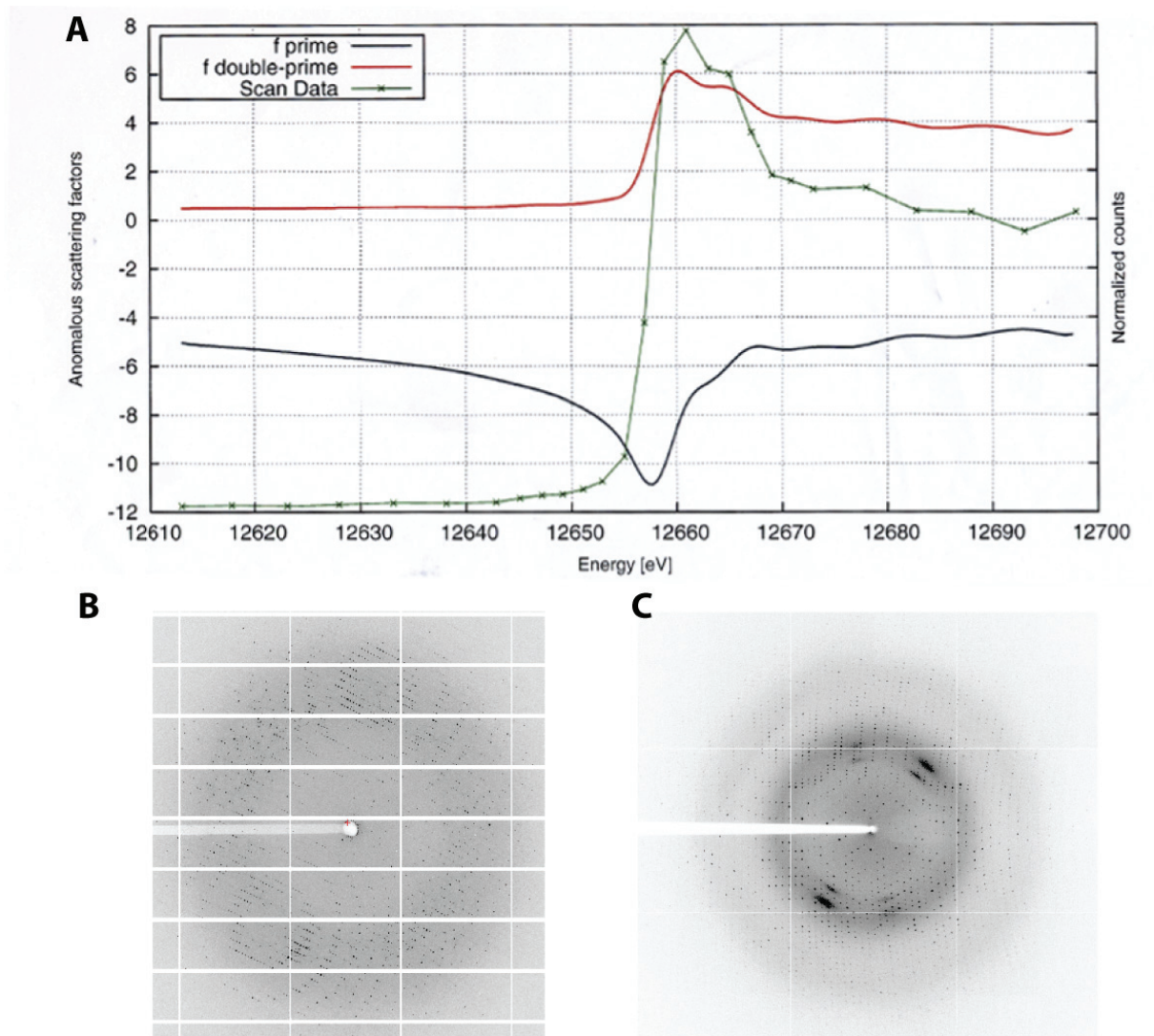


Figure 37: Data collection of GFP-*Enhancer* and GFP-*Minimizer* crystals.

(A) Fluorescence scan of GFP-*Enhancer* crystals at the K (1s) absorption edge of selene. (B) Diffraction image of GFP-*Enhancer* crystals and (C) GFP-*Minimizer* crystals.

8.4 Structure Determination and Refinement of the GFP-Enhancer and GFP-Minimizer Complexes

Table 14: Crystallographic Table for the GFP-Enhancer-Complex

	Selenomethionine	Native
Data collection	peak	
Space group	P4 ₂ 22	P6 ₁ 22
Cell dimensions		
a, b, c (Å)	a= 160.5 b= 160.5 c= 78.8	a= 87.4 b= 87.4 c=393.8
α, β, γ (°)	90, 90, 90	90, 90, 120
Wavelength	0.9793	1.00
Resolution (Å)	2.15	2.9
R _{sym}	3.9 (21.8)	4.7 (31.9)
I / σ I	14.35 (2.25)	20.93 (4.49)
Completeness (%)	99.3 (989)	99.3 (97.2)
Redundancy	3.13	3.85
<hr/>		
Refinement		
Resolution range (Å)	46.00-2.15	30.00-2.9
No. reflections	56271	31056
R _{work} / R _{free}	21.3. / 25.5	31.2 / 35.0
No. atoms		
Protein	5385	not specified
Water	407	not specified
B-factors		
Protein (Å ²)	46.5	76.95
Solvent (Å ²)	48.4	not specified
R.m.s deviations		
Bond lengths (Å)	0.008	0.009
Bond angles (°)	1.14	1.677

The native GFP-Enhancer dataset was indexed and scaled with a hexagonal bravais lattice. The rotation function indicated a six-fold symmetry. By analyzing the systemic absences an additional screw axis was identified. However, discrimination between the two enantiomorphic space groups P6₅22 and P6₁22 was impossible at this stage. Therefore the molecular replacement was performed with both space groups using the GFP structure (PDB ID 1GFL) and a llama VHH structure (PDB ID 1I3V) as two independent search assemblies (Yang et al. 1996; Spinelli et al. 2001). Only the replacement with space group P6₁22 resulted in a unique solution. Here, two GFP-Enhancer complexes were found in the asymmetric unit. Despite intensive efforts, employing non-crystallographic symmetry, all available refinement methods and manual model building, the refinement got trapped at an insufficient R factor (R_{work} 31.2, R_{free} 35.0). In consequence of the observed difficulties during

refinement, the obtained structure probably has a very strong model bias, which had not been overcome. Alternatively, the data sets may have suffered from crystal twinning. The statistics of the preliminary refinement can be found in **Table 14**.

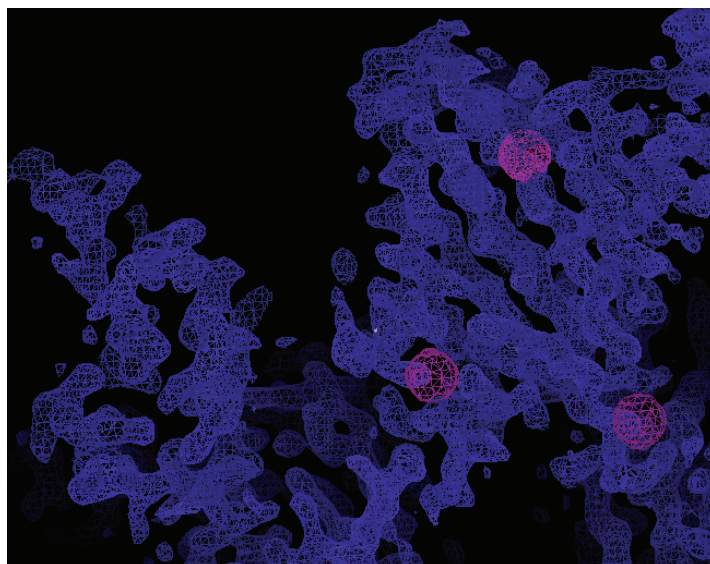
The peak data set collected from a single selenomethionine containing GFP-*Enhancer* crystal was indexed and scaled with space group $P4_22_2$. Three of four selenium sites in GFP could be located and a continuous electron density was generated after single anomalous wavelength dispersion (SAD) phasing and solvent flattening (**Figure 38**). Automatic model building was followed by manual modifications and iterative refinement steps which resulted in a final model with good stereochemistry and an R factor of 21.3. (R_{free} 25.5). Refinement statistics are listed in **Table 14**. The high-resolution data set of GFP-*Minimizer* crystals was integrated assuming $P2_12_12_1$ symmetry. Molecular replacement using the previously determined structure of GFP and *Enhancer* as independent search models resulted in a unique solution. The model was manually altered and gradually refined, resulting in a final model with an R factor of 16.1 (R_{free} 19.4). **Table 15** summarizes the refinement statistics.

8.5 Structure of GFP-*Enhancer* and GFP-*Minimizer* Complexes

To reveal the molecular mechanism underlying the observed fluorescence modulation crystal structures of the GFP-*Enhancer* and GFP-*Minimizer* complex were determined to 2.15 and 1.6 Å resolution, respectively. Strikingly, both nanobodies recognize two different, slightly overlapping epitopes on the GFP surface (**Figure 39A** and **B**). Thus, the observed competition for binding seems to result from a sterical clash between the nanobodies. *Enhancer* binds wtGFP in a front-wise manner at an exposed loop region between GFP β -strand 6 and 7 as well as parts of β -strand 8, making specific contacts with all three complementarity determining regions (CDRs) of the nanobody (**Figure 39A**). Previous structural studies of nanobodies have shown that the CDR3 normally folds over the framework-2 region that in the case of classical antibodies binds to the variable domain of the light-chain (VL). In contrast, the extremely short CDR3 of *Enhancer* is stretched out, thereby making the framework-2 region solvent-accessible in the antigen-free form. Surprisingly, this entire framework-2 area participates in GFP-recognition, contrasting the structure of classical antibodies where it would contact the VL-domain. Interestingly, the majority of the specific contacts are formed between CDR3 and GFP, whereas CDR1 and 2 remain solvent exposed. Notably, the interaction between GFP and *Enhancer* is predominantly electrostatic, spanning an interface of 672 Å² (**Table 16**). An additional non-polar contact is mediated by Phe98^{*Enhancer*}, which binds a hydrophobic surface patch on GFP, formed by Ala206^{GFP}, Leu221^{GFP} and Phe223^{GFP}.

Table 15: Crystallographic Table for the GFP-*Minimizer* Complex

Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
a, b, c (Å)	a= 50.78, b= 81.62, c= 94.48
Wavelength	0.98137
Resolution (Å)	1.5
R_{sym}	5.7 (39.1)*
$I / \sigma I$	16.91 (2.84)
Completeness (%)	97.5 (86.3)
Redundancy	3.85
Refinement	
Resolution range (Å)	47.00-1.61
No. reflections	49989
$R_{\text{work}} / R_{\text{free}}$	16.1 / 19.4
No. atoms	
Protein	2818
Water	685
<i>B</i> -factors	
Protein (Å ²)	13.7
Solvent (Å ²)	30.1
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.04

**Figure 38:** Experimentally determined electron density of the GFP-*Enhancer* complex.

Multiple wavelength anomalous dispersion map of the GFP Enhancer complex at 2.5 Å resolution after solvent flattening. Selenide sites are shown as magenta spheres. The characteristic β -barrel fold of GFP can be easily identified.

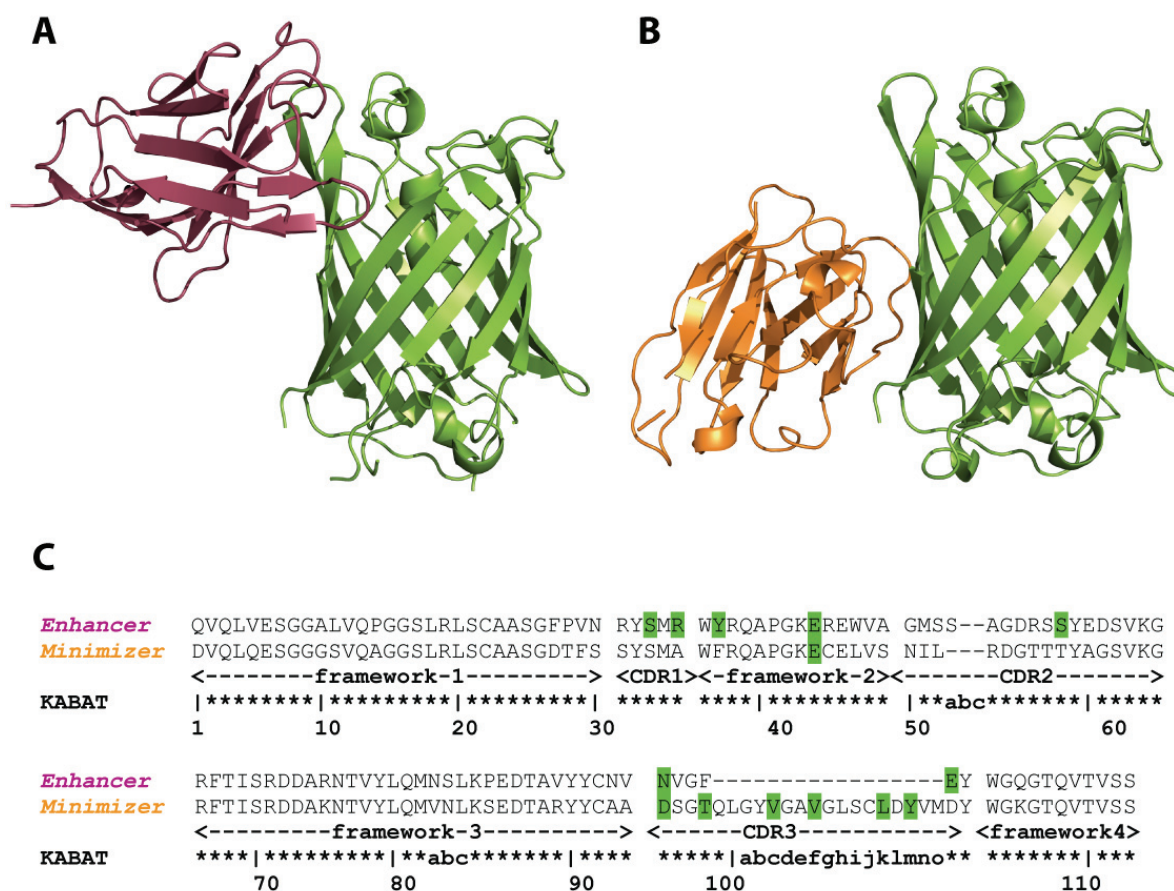


Figure 39: Structural overview of GFP nanobody complexes.

Overall structure of the (A) GFP-*Enhancer* and (B) the GFP-*Minimizer* complex (ribbon presentation). The GFP molecules are colored in green and the nanobodies in raspberry (*Enhancer*) and orange (*Minimizer*), respectively. (C) Sequence alignment of nanobodies with GFP contacting residues colored in green. Framework regions and complementary determining regions (CDRs) are annotated. Residue numbering is according to Kabat (Kabat and Wu 1991).

The *Minimizer* binds GFP in a sidewise orientation, employing its elongated CDR3 to target β -strand 6 and 7 of GFP (**Figure 39B**). The interaction with GFP is quite remarkable, since the nanobody targets the rigid and flat side rather than the more flexible and easily accessible loops at the top and the bottom of the β -can. The *Minimizer* occludes a slightly smaller surface area (652 \AA^2) on GFP and the overall number of contacts is less, in line with its ready replacement by the *Enhancer* in the competition assay (**Figure 39C**, **Table 16**).

Table 16: Salt bridges and hydrogen bonds between GFP and *Enhancer*

Salt bridges and hydrogen bonds				
GFP Residue	GFP Atom	<i>Enhancer</i> Residue	<i>Enhancer</i> Atom	Distance [Å]
Glu142	OE2	Ser33	OG	2.42
Glu142	OE1	Ser33	OG	2.96
Glu142	OE2	Arg35	NH2	2.54
Tyr145	O	Asn95	ND2	2.93
Asn 146	OD1	Asn95	ND2	3.07
Ser147	N	Glu101	OE2	2.79
Lys166	NZ	Glu44	OE1	2.61
Arg168	NH1	Glu101	OE2	3.19
Arg168	NH2	Glu101	OE1	2.82
Arg168	NH2	Tyr37	OH	3.25
Asp173	O	Ser58	OG	2.88
Gly174	O	Arg35	NH1	3.08
Ser175	O	Arg35	NH1	2.42
Hydrophobic interactions				
GFP Residue	<i>Enhancer</i> Residue			
Ala206	Phe98			
Leu221	Phe98			
Phe223	Phe98			

8.6 Binding of Nanobodies Induces Structural Rearrangements in the Chromophore Environment

In general, binding of the nanobodies has no substantial global influence on the overall fold of GFP. wtGFP (PDB ID 1EMB) has RMSDs of 0.310 Å ($C\alpha$ atoms) and 0.357 Å ($C\alpha$ atoms) to *Enhancer* and to *Minimizer*, respectively. From previous structural studies on GFP, however, it is known that already minimal perturbations in the chromophore environment can have vast effects on its fluorescence properties (Brejc et al. 1997). Indeed, a detailed comparison of the GFP-nanobody structures with previously published GFP structures reveals, that the GFP-*Enhancer* complex harbors the deprotonated, negatively charged state of the GFP-chromophore which has been described for GFP(S65T) (Brejc et al. 1997) (**Figure 40A**). Binding of *Enhancer* induces slight structural shifts in the loop region from amino acid Glu142^{GFP} to His148^{GFP} and fixes Arg168^{GFP} in close proximity to His148^{GFP}. The conformation of the Arg168^{GFP} side chain is stabilized by direct contacts with *Enhancer*

residues Tyr37^{Enhancer} and Glu101^{Enhancer}. These structural rearrangements bring the proton donor His148^{GFP} in very close proximity to the hydroxyl group of the GFP chromophore (distance 2.7 Å compared to 2.8 Å for GFP(S65T) and 3.4 Å for wtGFP). Thus, binding of the *Enhancer* likely facilitates improved proton extraction from the fluorophore hydroxyl by His148^{GFP}, thereby stabilizing the phenolate anion of the fluorophore and enhancing the fluorescence intensity.

Table 17: Salt bridges and hydrogen bonds between GFP and *Minimizer*

GFP Residue	GFP Atom	Minimizer Residue	Minimizer Atom	Distance [Å]
Asn 149	ND2	Val 100f	O	3.12
Tyr 151	OH	Val 100f	O	2.63
Tyr 151	OH	Asp100l	O	3.19
Asn 164	ND2	Tyr 100m	OH	3.09
Lys 166	N2	Asp 95	OD1	2.70
Arg 168	NH2	Leu 100k	O	2.97
Asp 180	OD1	Thr 98	N	2.98
Asp 180	OD2	Thr 98	OG1	2.66
Asn 198	ND2	Val 100c	O	3.05
Tyr 200	OH	Glu 44	OE1	2.64
Asn 149	ND2	Val 100f	O	3.12
Tyr 151	OH	Val 100f	O	2.63

In contrast, the chromophore environment of the GFP-*Minimizer* complex is remarkably different and shows similarities to the situation present in the less-fluorescent wtGFP (Brejc et al. 1997). Of note here, Arg168^{GFP} is rather flexible in comparison to the *Enhancer* complex (**Figure 40B**). Two alternative conformations of its guanidine group could be traced in the electron density. In one of the conformations Arg168^{GFP} is tilted away from His148^{GFP} and instead makes specific contacts to the backbone carbonyl of Leu100i^{Minimizer}. This nanobody-induced conformational modulation reduces the electrostatic forces exerted on His148^{GFP}, which is in fact pulled back from the hydroxyl group of the chromophore and positioned with 3.5 Å distance (wtGFP: 3.4 Å), too far to efficiently stabilize the phenolate anion. Instead, binding of *Minimizer* likely stabilizes an arrangement of the chromophore-surrounding environment, which favors the neutral phenol state of the chromophore. In support of this model, the fluorescence enhancement of eGFP - where the phenolate anion state is stabilized by an engineered mutation - by the *Enhancer* is only 1.5 fold compared to 5 fold for wtGFP, while the fluorescence suppression is 8 fold, compared to 4 fold for wtGFP.

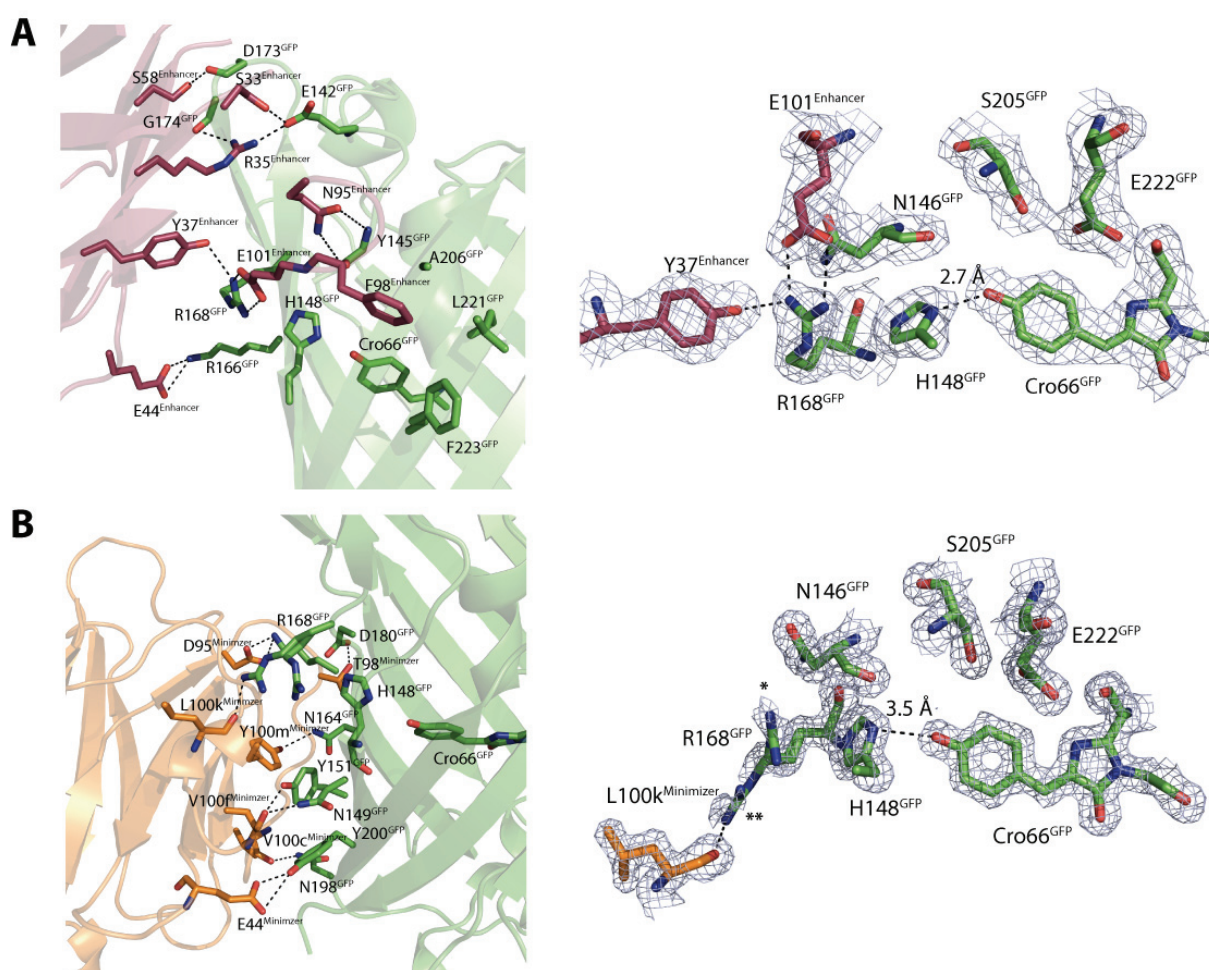


Figure 40: Nanobody binding to GFP and induced structural rearrangements in the chromophore environment.

An enlargement of parts of the binding sites with crucial residues and the GFP chromophore (Cro66) highlighted as sticks. The panels on the right show the GFP chromophore environment for the GFP-*Enhancer* (A) and the GFP-*Minimizer* complex (B), respectively, plus a 2.15 Å and a 1.6 Å electron density map calculated with 2 $f_0 - f_c$ coefficients and contours at 1.0 σ . Binding of *Enhancer* (A) induces slight structural rearrangements which bring the proton donor His148 in closer proximity to the hydroxyl group of chromophore. R168^{GFP} is stabilized by direct contacts to Y37^{Enhancer} and E101^{Enhancer}. In the case of *Minimizer* binding (B), the guanine group of R168^{GFP} is rather flexible. Two alternative conformations for its side chain (marked with * and **) were traced. One conformation is tilted away from H148^{GFP} making specific contacts to L100i^{Minimizer}.

In summary, the two nanobodies appear to stabilize and by thermodynamic reasoning preferentially recognize remarkably subtle conformational states of GFP, inflicted by an internal protonation / deprotonation reaction. Such a capability to distinguish small functional states could be an important tool in the future, also in the context of modulating the properties of other proteins.

8.7 The *Enhancer* and the *Minimizer* Interfere with GFP Chromophore Environment

To further test the structure-derived hypothesis that the interactions of the two nanobodies stabilize the neutral (*Minimizer*) or the ionized state of the chromophore (*Enhancer*), fluorescence absorption spectra were analyzed (**Figure 41A and B**). In support of this model, the *Enhancer* increases absorption at 475 nm (wtGFP: 4 fold increase, eGFP: 1.5 fold increase) while reducing it at 395nm for both wtGFP and eGFP (wtGFP 1.2 fold decrease , eGFP: 3 fold decrease). The *Minimizer* modulates the absorption in an exactly opposite manner, reducing absorption at 475 nm (wtGFP: 1.5 decrease, eGFP: 5 fold decrease) and increasing it a 395 nm (wtGFP 1.4 fold increase; eGFP: 1.8 fold increase). Therefore, the readily observed fluorescence modulation upon nanobody binding is presumably due to a change in the absorption efficiency at different wavelengths, which correlates with the magnitude of fluorescence emission and can be attributed to the protonation state of the fluorophore. The nanobody induced structural changes in the chromophore environment are summarized in **Figure 41C**.

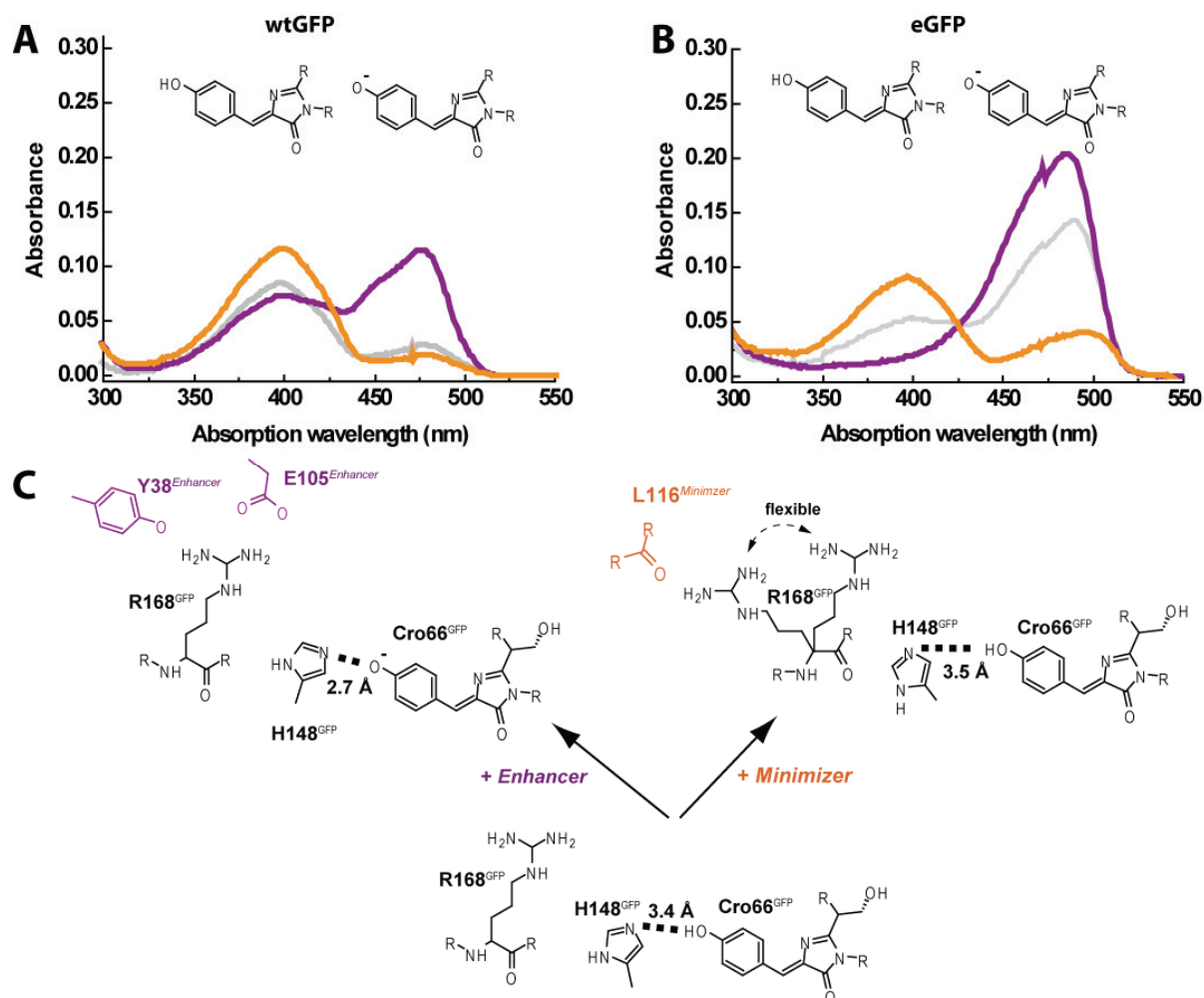


Figure 41: Nanobody-induced GFP spectroscopic changes.

Absorption spectra of the unbound (grey) *Minimizer*-bound (orange) or *Enhancer*-bound (raspberry) wtGFP (A) or eGFP (B). The absorption at 395 nm corresponds to the protonated chromophore, absorption at 475 nm to the anionic chromophore (see chemical notation above). (C) Schematic illustration of the nanobody-induced structural changes in the GFP chromophore environment. Distances between the chromophore and H148^{GFP} are indicated.

9 Discussion

9.1 *Enhancer* and *Minimizer* Modulate GFP Intensity in Living Cells

By screening a selection of GFP binding proteins, two nanobodies were found, that reversibly modulate GFP fluorescence; one nanobody minimized GFP fluorescence about fivefold and displacement by a second nanobody caused a tenfold increase. Structural analysis of GFP-nanobody complexes revealed that the two nanobodies induce and stabilize subtle opposing changes in the chromophore environment leading to drastically altered absorption properties. Since GFP is widely used as fluorescence marker in cell imaging, an application of the newly discovered phenomenon is highly desirable. The following paragraph describes the nanobody-induced modulation in living cells. These experiments were conducted in the laboratory of Prof. Leonhardt and Dr. Ulrich Rothbauer, in great parts by Jonas Helma.

It was tested whether the nanobody-induced fluorescence modulation observed *in vitro* also occurs *in vivo*. To this end, human embryonic kidney (HEK) 293T cells with expression vectors encoding wtGFP or eGFP in combination were transfected with constructs encoding *Enhancer*, *Minimizer* or a *Control* nanobody fused to monomeric red fluorescent protein (mRFP). After two days of incubation combined excitation and emission scans of GFP fluorescence intensities were performed in living cells (**Figure 42A** and **B**). Indeed, both nanobodies-induced spectral changes of wtGFP and eGFP fluorescence *in vivo* were similar to the ones observed *in vitro* and described above. Co-expression of the *Enhancer* and its target increased the signal of wtGFP at least 2-fold, whereas co-expression of the *Minimizer* resulted in a more than 5-fold decrease (**Figure 42C**). The effects of the nanobodies on eGFP were not as drastic but show a similar tendency. Altogether, these data clearly demonstrate that *Enhancer* as well as *Minimizer* can be effectively applied to induce a fluorescence modulation of GFP in living cells.

To obtain a concentration-independent measurement of *Enhancer* or *Minimizer* binding and to test whether subcellular differences can be detected, the induced shift of GFP absorption maxima from 405 nm to 488 nm was determined by ratio imaging. To this end, the *Enhancer* was tethered to the nuclear lamina by transfecting HeLa cells with an expression construct coding for an *Enhancer*-Lamin B1 fusion. After co-expressing wtGFP in excess, images were acquired with excitation at 405 nm and at 488 nm to detect relative differences in GFP fluorescence intensities at the nuclear lamina due to binding to locally immobilized *Enhancer*-Lamin B1 nanobody fusion protein. Although GFP was bound and enriched at the nuclear

lamina, this structure was barely detectable after excitation at 405 nm, however, excitation at 488 nm indeed lead to an increased signal at the nuclear lamina (**Figure 42D**). Subsequently, the presence of *Enhancer* at this distinct sub-cellular structure could be visualized by calculating the ratio between pixel intensities obtain at 405 nm and at 488 nm, displaying the ratio images in false color, thereby distinguishing bound from unbound GFP (**Figure 42D**). This example illustrates how fluorescence-modulating nanobodies enable novel optical readouts.

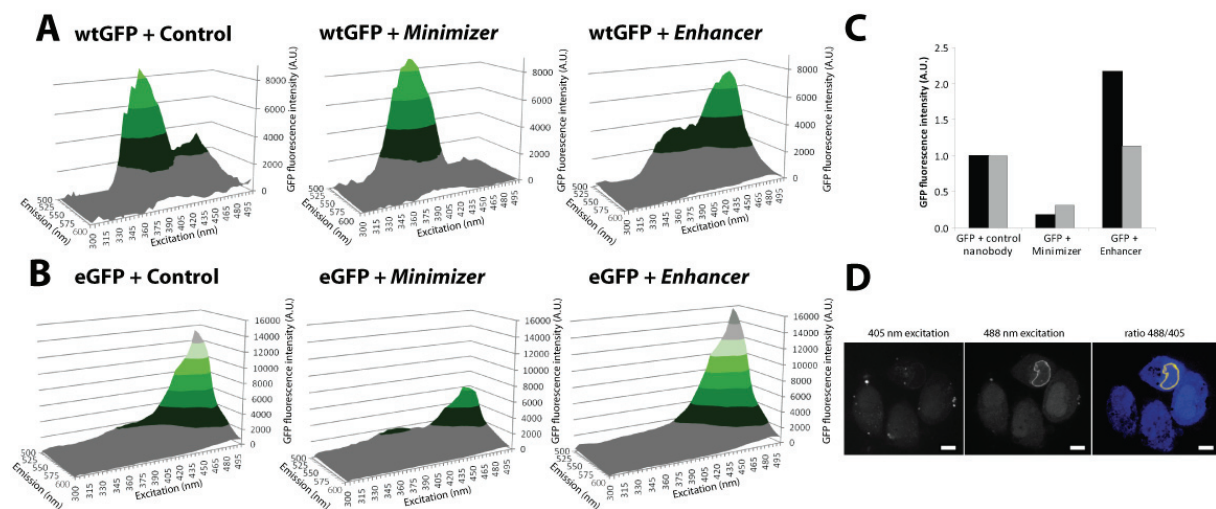


Figure 42: The nanobody-induced fluorescence modulation is applicable in living cells.

Binding of *Minimizer* and *Enhancer* shifts excitation and emission spectra for both, wtGFP (A) and eGFP (B) *in vivo*. (C) Enhancement and minimization factors for wtGFP (black bars) and eGFP (grey bars) fluorescence were calculated by comparing overall fluorescence intensities to expression levels *in vivo*. After coexpression of *Minimizer* a significant reduction of the fluorescence intensity of wtGFP as well as eGFP is observable. The presence of the *Enhancer* increases the fluorescence intensity of both wtGFP and eGFP. Values obtained for the control nanobody were set to 1. (D) Ratio imaging. Shown are cells expressing wtGFP, which is dispersedly distributed throughout the cell. The top-most cell coexpresses *Enhancer* fused to laminB1 (GBP1-Lamin-B1). While only a weak signal at the nuclear lamina is detectable with excitation at 405 nm the relative and absolute signal increased with excitation at 488 nm due to the absorption modulation induced by binding to the *Enhancer*. Bound and unbound GFP can be distinguished independent of the total concentration by matrix algebra, calculating the ratio between the signal intensities obtained with excitation at 488 nm and at 405 nm for every pixel. The 488/405 ratios are then displayed in a false color gradient from blue (bound) to yellow (unbound). Scale bar is 10 μ m.

These ratio-imaging experiments also served as an initial proof of principle, that the nanobodies were applicable *in vivo*.

Next, it should be tested whether the fluorescence enhancement effect induced by binding of GFP to *Enhancer* localized in a defined subcellular compartment, could be used to track subcellular translocation events in a high throughput manner. As an example, the inducible

translocation of the human estrogen receptor (ER) was examined. In this system, hormone binding leads to a conformational change in the receptor that results in its dissociation from chaperone proteins and ultimately in the binding of the receptor as a homodimer to cognate sites in steroid-responsive genes in the nucleus (Kabat and Wu 1991; Tsai and O'Malley 1994). This subcellular trafficking event can be induced by the synthetic steroid hormone tamoxifen and followed with a GFP-labeled receptor and high resolution fluorescence microscopy (Htun et al. 1999). Since this original procedure is based on single cell imaging it is poorly suited for high throughput analysis. However, it was reasoned that the fixed-*Enhancer* tracking system would allow for a simple fluorescent readout, thus making a highthroughput setup applicable. To test our hypothesis, a mammalian (HeLa-Kyoto) cell line was generated that stably expresses nuclear localized *Enhancer* fused to mRFP (nls-*Enhancer*). The newly constructed cell line was utilized to transiently co-express the steroid binding domain of the human estrogen receptor (ER₂₈₆₋₅₉₅) fused to wtGFP (GFP-ER₂₈₆₋₅₉₅). According to our prediction, translocation of this construct from the cytoplasm to the nucleus should be detectable by an increase of the GFP-fluorescence intensity upon binding of GFP to the nls-*Enhancer* in the nucleus (**Figure 43A**). Using fluorescence microscopy it was confirmed, that both constructs - GFP-ER₂₈₆₋₅₉₅ and nls-*Enhancer* - are almost exclusively targeted to their designated compartment. Upon addition of tamoxifen to the medium, GFP-ER₂₈₆₋₅₉₅ inducibly translocates from the cytoplasm into the nucleus (**Figure 43B**). Entering the nucleus, GFP becomes accessible for binding to nls-*Enhancer* which indeed results in a threefold increase in GFP-fluorescence intensity (**Figure 43C**). Importantly, the translocation event can be followed in a statistically significant number of cells by scanning the fluorescence intensities of living cells in multi-well formats. Moreover the fluorescence enhancement is directly correlated with translocation efficiency. By quantifying the fluorescence intensities a clear dose-dependency of translocation upon addition of increasing amounts of tamoxifen was detected (**Figure 43D**). Moreover, since our assay uses living cells, it was possible to follow the dynamics of the translocation event over time. These data demonstrate that fluorescence modulating nanobodies are powerful tools to study subcellular re-localization, a key process of many signal transduction pathways, in real time, high throughput and in a quantitative manner.

9.2 Modulation of GFP Fluorescence with Nanobodies – Future Perspectives

The modulation of GFP fluorescence enables a number of new bioimaging applications. The expression and subcellular distribution of *Minimizer* and *Enhancer* can be detected by ratio

imaging, allowing the distinction of bound from unbound GFP. This indirect optical readout can be used as reporter for gene expression, virus infection and translocation assays. The nanobody-mediated enhancement of GFP fluorescence should also improve the tracing of low-abundant GFP fusion proteins in live cells as well as ultrahigh resolution microscopy. Recently, it was demonstrated that cellular structures can be imaged at sub-diffraction resolution by 3D structured illumination microscopy (3D-SIM, (Schermelleh et al. 2008)). To obtain ultrahigh resolution, this new microscopy technology, however, requires hundreds of images and thus mostly relies on bright synthetic chromophores. The signals obtained with GFP labeled proteins are barely sufficient. Therefore, any fluorescence enhancement by either co-expressed or subsequently added *Enhancer* nanobody would greatly facilitate 3D-SIM. Translocation events play a central role in signal transduction and are therefore a prime target for drug screenings. Presently, translocations are either indirectly monitored with reporter gene assays that take at least a day or by microscopy which requires costly and technically demanding high throughput image acquisition and analysis tools. Our nanobody based assay can be performed with a simple plate reader and measures translocation as fluorescence enhancement after drug addition. The feasibility of this assay principle was demonstrated using the tamoxifen-induced nuclear translocation of the estrogen receptor. Besides steroid hormone receptors also notch type signaling in differentiation and cancer and presenilin / gamma secretase activity in Alzheimer's disease could be directly monitored in cell based drug screens.

This work also opens new perspectives as it exemplarily shows applications of nanobodies ranging from affinity purification and crystallization of proteins to the manipulation of conformational states and protein function *in vitro* and *in vivo*. Especially the detection and manipulation of alternative protein conformation in living cells enable novel types of studies in molecular and cellular biology. The functional relevance of alternative protein conformations is clearly illustrated by the prion protein. Here an experimental strategy to address the role of alternative protein conformations in cellular systems is demonstrated. The here presented results show that nanobodies can be generated to recognize, induce and stabilize alternative protein conformations and thus enable studies of functional properties of specific protein conformations *in vitro* and *in vivo*.

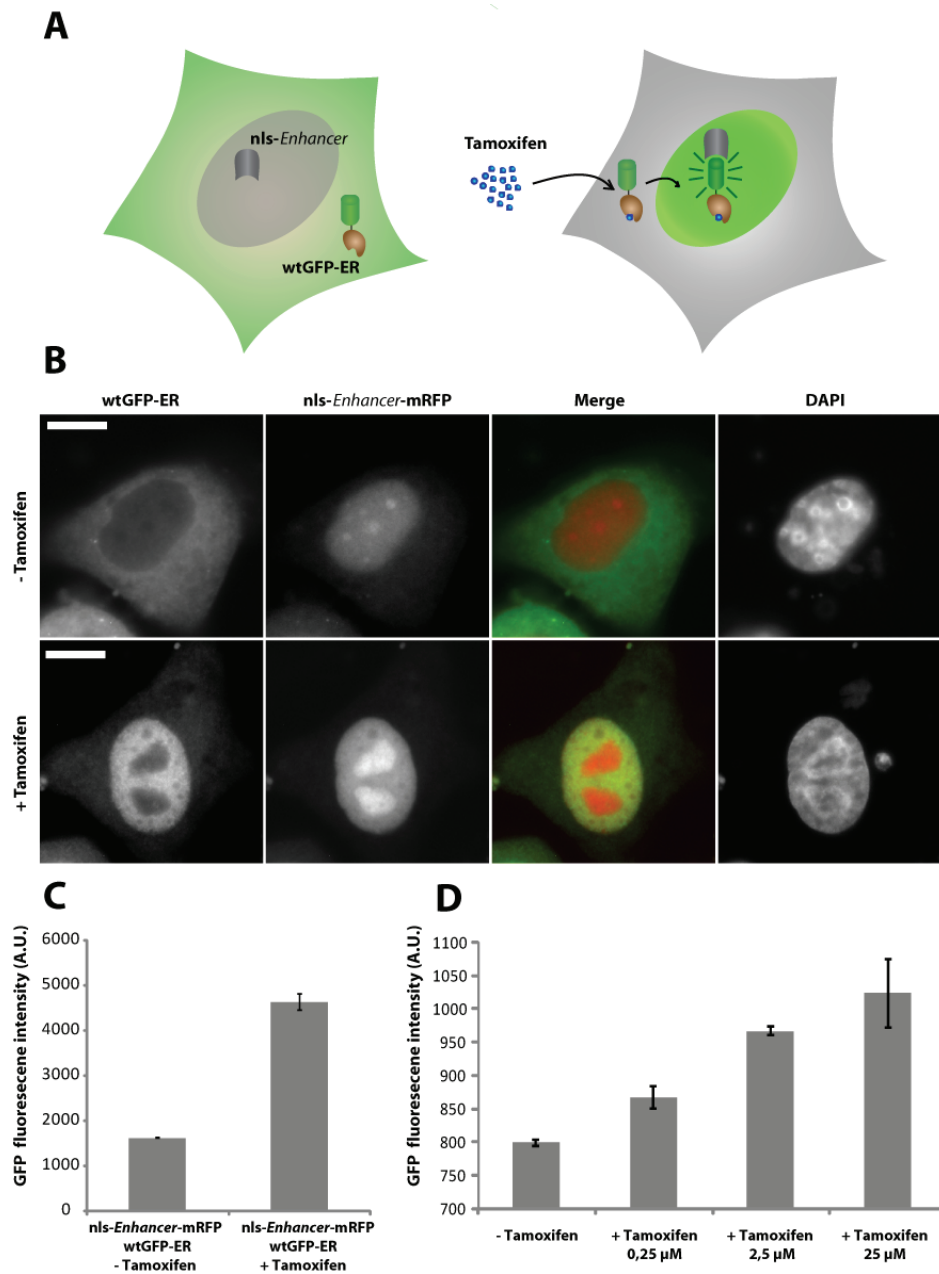


Figure 43: Cytoplasmic-nuclear translocation events can be detected by nanobody mediated fluorescence enhancement.

(A) Schematic outline of the translocation assay: GFP-ER and nls-*Enhancer* are present in the cytoplasm and the nucleus, respectively. Addition of Tamoxifen induces translocation of GFP-ER to the nucleus where it binds to nls-*Enhancer* leading to an increase of the fluorescence intensity. (B) Representative cells were analyzed by fluorescence microscopy (scale bar 10 μm). Untreated cells (-Tamoxifen, upper row) show an almost exclusive distribution of GFP-ER to the cytoplasm while the nls-*Enhancer* is localized in the nucleus. After addition of Tamoxifen (+Tamoxifen, lower row) GFP-ER co-localizes with nls-*Enhancer* in the nucleus. (C, D). Nucleo-cytoplasmic translocation induced by Tamoxifen was assayed in a plate format in living cells. (C) After translocation of GFP-ER into the nucleus binding of nls-*Enhancer* leads to a 3-fold increase of the fluorescence intensity. (D) Detection of dose dependent translocation efficiency by measuring the fluorescence enhancement effect. Increasing concentrations of Tamoxifen are indicated.

10 Summary

The development of the green fluorescent protein (GFP) and its spectral variants was a milestone-discovery and transformed cell biological research. Although a range of GFP variants have been engineered to date, the possibility to alter GFP fluorescence is still lacking. By screening a selection of GFP binding nanobodies two candidates were found, which were termed *Minimizer* and *Enhancer* nanobody, respectively. Remarkably, binding of these two nanobodies antagonistically influenced the fluorescence properties of GFP, constituting a reversible fluorescence switch. The *Minimizer* induced a 5-fold diminishment of GFP emission intensity which can be reversed by addition of *Enhancer* leading to an overall 20-fold increase of fluorescence. Structural and spectroscopic analysis of the two GFP-nanobody complexes revealed that only subtle changes in the chromophore environment were responsible for the drastically altered absorption properties. Furthermore, it was demonstrated that both nanobodies can modulate the fluorescence intensity of GFP-fusions in living cells. The nanobody induced fluorescence modulation could be detected by ratio imaging and used to measure expression and subcellular localization. Furthermore, the fluorescence modulation can be employed as cellular translocation assay. Cells expressing nuclear nanobodies showed increased fluorescence upon tamoxifen-induced translocation of GFP-labeled estrogen receptor, providing a simple readout for functional studies and drug screenings. This work demonstrates that alternative protein conformations can be manipulated and studied with nanobodies, making these binders a novel and attractive tool for future research.

11 References

- Adams, P. D., R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter and T. C. Terwilliger (2002). "PHENIX: building new software for automated crystallographic structure determination." Acta Crystallogr D Biol Crystallogr **58**(Pt 11): 1948-54.
- Ambros, V. and D. Baltimore (1978). "Protein is linked to the 5' end of poliovirus RNA by a phosphodiester linkage to tyrosine." J Biol Chem **253**(15): 5263-6.
- Andrejeva, J., K. S. Childs, D. F. Young, T. S. Carlos, N. Stock, S. Goodbourn and R. E. Randall (2004). "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 **101**(49): 17264-9.
- Areschoug, T. and S. Gordon (2009). "Scavenger receptors: role in innate immunity and microbial pathogenesis." Cell Microbiol.
- Arimoto, K., H. Konishi and K. Shimotohno (2008). "UbcH8 regulates ubiquitin and ISG15 conjugation to RIG-I." Mol Immunol **45**(4): 1078-84.
- Arimoto, K., H. Takahashi, T. Hishiki, H. Konishi, T. Fujita and K. Shimotohno (2007). "Negative regulation of the RIG-I signaling by the ubiquitin ligase RNF125." Proc Natl Acad Sci U S A **104**(18): 7500-5.
- Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-41.
- Berger, I., D. J. Fitzgerald and T. J. Richmond (2004). "Baculovirus expression system for heterologous multiprotein complexes." Nat Biotechnol **22**(12): 1583-7.
- Biegert, A., C. Mayer, M. Remmert, J. Soding and A. N. Lupas (2006). "The MPI Bioinformatics Toolkit for protein sequence analysis." Nucleic Acids Res **34**(Web Server issue): W335-9.
- Biocca, S., A. Di Luzio, T. Werge and A. Cattaneo (1991). "Intracellular immunization: expression of antibody domains in the cytoplasm and in the nucleus of mammalian cells." Cytotechnology **5 Suppl 1**: 49-50.
- Bowie, A. G. and K. A. Fitzgerald (2007). "RIG-I: tri-ning to discriminate between self and non-self RNA." Trends Immunol **28**(4): 147-50.

- Brejč, K., T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormo and S. J. Remington (1997). "Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein." Proc Natl Acad Sci U S A **94**(6): 2306-11.
- Cattaneo, A. and S. Biocca (1999). "The selection of intracellular antibodies." Trends Biotechnol **17**(3): 115-21.
- Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher (1994). "Green fluorescent protein as a marker for gene expression." Science **263**(5148): 802-5.
- Chattoraj, M., B. A. King, G. U. Bublitz and S. G. Boxer (1996). "Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer." Proc Natl Acad Sci U S A **93**(16): 8362-7.
- Childs, K., N. Stock, C. Ross, J. Andrejeva, L. Hilton, M. Skinner, R. Randall and S. Goodbourn (2007). "mda-5, but not RIG-I, is a common target for paramyxovirus V proteins." Virology **359**(1): 190-200.
- Childs, K. S., J. Andrejeva, R. E. Randall and S. Goodbourn (2008). "Mechanism of mda-5 inhibition by paramyxovirus V proteins." J Virol.
- Claude, J. B., K. Suhre, C. Notredame, J. M. Claverie and C. Abergel (2004). "CaspR: a web server for automated molecular replacement using homology modelling." Nucleic Acids Res **32**(Web Server issue): W606-9.
- Colby, C. and M. J. Chamberlin (1969). "The specificity of interferon induction in chick embryo cells by helical RNA." Proc Natl Acad Sci U S A **63**(1): 160-7.
- Cormier, M. J., K. Hori, Y. D. Karkhanis, J. M. Anderson, J. E. Wampler, J. G. Morin and J. W. Hastings (1973). "Evidence for similar biochemical requirements for bioluminescence among the coelenterates." J Cell Physiol **81**(2): 291-7.
- Cortez-Retamozo, V., N. Backmann, P. D. Senter, U. Wernery, P. De Baetselier, S. Muyldermans and H. Revets (2004). "Efficient cancer therapy with a nanobody-based conjugate." Cancer Res **64**(8): 2853-7.
- Cui, S., K. Eisenacher, A. Kirchhofer, K. Brzozka, A. Lammens, K. Lammens, T. Fujita, K. K. Conzelmann, A. Krug and K. P. Hopfner (2008). "The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I." Mol Cell **29**(2): 169-79.
- De Genst, E., K. Silence, K. Decanniere, K. Conrath, R. Loris, J. Kinne, S. Muyldermans and L. Wyns (2006). "Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies." Proc Natl Acad Sci U S A **103**(12): 4586-91.

- Desmyter, A., S. Spinelli, F. Payan, M. Lauwereys, L. Wyns, S. Muyldermans and C. Cambillau (2002). "Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology." J Biol Chem **277**(26): 23645-50.
- Desmyter, A., T. R. Transue, M. A. Ghahroudi, M. H. Thi, F. Poortmans, R. Hamers, S. Muyldermans and L. Wyns (1996). "Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme." Nat Struct Biol **3**(9): 803-11.
- Drenth, J. (2002). "Principles of Protein X-Ray Crystallography." Springer Second Edition.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.
- Eswar, N., D. Eramian, B. Webb, M. Y. Shen and A. Sali (2008). "Protein structure modeling with MODELLER." Methods Mol Biol **426**: 145-59.
- Faustin, B., L. Lartigue, J. M. Bruey, F. Luciano, E. Sergienko, B. Bailly-Maitre, N. Volkmann, D. Hanein, I. Rouiller and J. C. Reed (2007). "Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase-1 activation." Mol Cell **25**(5): 713-24.
- Gack, M. U., A. Kirchhofer, Y. C. Shin, K. S. Inn, C. Liang, S. Cui, S. Myong, T. Ha, K. P. Hopfner and J. U. Jung (2008). "Roles of RIG-I N-terminal tandem CARD and splice variant in TRIM25-mediated antiviral signal transduction." Proc Natl Acad Sci U S A **105**(43): 16743-8.
- Gack, M. U., Y. C. Shin, C. H. Joo, T. Urano, C. Liang, L. Sun, O. Takeuchi, S. Akira, Z. Chen, S. Inoue and J. U. Jung (2007). "TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity." Nature **446**(7138): 916-920.
- Gay, N. J. and M. Gangloff (2007). "Structure and function of Toll receptors and their ligands." Annu Rev Biochem **76**: 141-65.
- Giacovazzo, C. (2003). "Fundamentals of Crystallography." Oxford University Press Second Edition.
- Guo, Z., L. M. Chen, H. Zeng, J. A. Gomez, J. Plowden, T. Fujita, J. M. Katz, R. O. Donis and S. Sambhara (2007). "NS1 protein of influenza A virus inhibits the function of intracytoplasmic pathogen sensor, RIG-I." Am J Respir Cell Mol Biol **36**(3): 263-9.
- Hamers-Casterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman and R. Hamers (1993). "Naturally occurring antibodies devoid of light chains." Nature **363**(6428): 446-8.

- Heim, R., A. B. Cubitt and R. Y. Tsien (1995). "Improved green fluorescence." Nature **373**(6516): 663-4.
- Heim, R., D. C. Prasher and R. Y. Tsien (1994). "Wavelength mutations and posttranslational autoxidation of green fluorescent protein." Proc Natl Acad Sci U S A **91**(26): 12501-4.
- Heim, R. and R. Y. Tsien (1996). "Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer." Curr Biol **6**(2): 178-82.
- Hendrickson, W. A., J. R. Horton and D. M. LeMaster (1990). "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure." Embo J **9**(5): 1665-72.
- Hornung, V., J. Ellegast, S. Kim, K. Brzozka, A. Jung, H. Kato, H. Poeck, S. Akira, K. K. Conzelmann, M. Schlee, S. Endres and G. Hartmann (2006). "5'-Triphosphate RNA is the ligand for RIG-I." Science **314**(5801): 994-7.
- Horvath, C. M. (2004). "Weapons of STAT destruction. Interferon evasion by paramyxovirus V protein." Eur J Biochem **271**(23-24): 4621-8.
- Htun, H., L. T. Holth, D. Walker, J. R. Davie and G. L. Hager (1999). "Direct visualization of the human estrogen receptor alpha reveals a role for ligand in the nuclear distribution of the receptor." Mol Biol Cell **10**(2): 471-86.
- Janeway, C. A., P. Travers and M. Walport (2004). "Immunobiology." **6**.
- Jin, M. S. and J. O. Lee (2008). "Structures of the toll-like receptor family and its ligand complexes." Immunity **29**(2): 182-91.
- Kabat, E. A. and T. T. Wu (1991). "Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites." J Immunol **147**(5): 1709-19.
- Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." J Appl Cryst **26**: 795-800.
- Kato, H., O. Takeuchi, E. Mikamo-Satoh, R. Hirai, T. Kawai, K. Matsushita, A. Hiiragi, T. S. Dermody, T. Fujita and S. Akira (2008). "Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5." J Exp Med **205**(7): 1601-10.

- Kato, H., O. Takeuchi, S. Sato, M. Yoneyama, M. Yamamoto, K. Matsui, S. Uematsu, A. Jung, T. Kawai, K. J. Ishii, O. Yamaguchi, K. Otsu, T. Tsujimura, C. S. Koh, C. Reis e Sousa, Y. Matsuura, T. Fujita and S. Akira (2006). "Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses." Nature **441**(7089): 101-5.
- Kaukinen, P., M. Sillanpaa, S. Kotenko, R. Lin, J. Hiscott, K. Melen and I. Julkunen (2006). "Hepatitis C virus NS2 and NS3/4A proteins are potent inhibitors of host cell cytokine/chemokine gene expression." Virology **3**: 66.
- Kawai, T. and S. Akira (2007). "TLR signaling." Semin Immunol **19**(1): 24-32.
- Kawai, T., K. Takahashi, S. Sato, C. Coban, H. Kumar, H. Kato, K. J. Ishii, O. Takeuchi and S. Akira (2005). "IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction." Nat Immunol **6**(10): 981-8.
- Komuro, A. and C. M. Horvath (2006). "RNA- and virus-independent inhibition of antiviral signaling by RNA helicase LGP2." J Virol **80**(24): 12332-42.
- Konarev, P. V., M. V. Petoukhov, V. V. Volkov and D. I. Svergun (2006). "ATSAS 2.1, a program package for small-angle scattering data analysis." J Appl Cryst **39**: 277-286.
- Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-97.
- Kunkel, T. A. (1985). "Rapid and efficient site-specific mutagenesis without phenotypic selection." Proc Natl Acad Sci U S A **82**(2): 488-92.
- Lauwereys, M., M. Arbabi Ghahroudi, A. Desmyter, J. Kinne, W. Holzer, E. De Genst, L. Wyns and S. Muyldermans (1998). "Potent enzyme inhibitors derived from dromedary heavy-chain antibodies." Embo J **17**(13): 3512-20.
- Li, T., X. Chen, K. C. Garbutt, P. Zhou and N. Zheng (2006). "Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase." Cell **124**(1): 105-17.
- Li, X., C. T. Ranjith-Kumar, M. T. Brooks, S. Dharmiah, A. B. Herr, C. Kao and P. Li (2009). "The RIG-I like receptor LGP2 recognizes the termini of double-stranded RNA." J Biol Chem.
- Malathi, K., B. Dong, M. Gale, Jr. and R. H. Silverman (2007). "Small self-RNA generated by RNase L amplifies antiviral innate immunity." Nature **448**(7155): 816-9.
- Manzan, A., G. Pfeiffer, M. L. Hefferin, C. E. Lang, J. P. Carney and K. P. Hopfner (2004). "MlaA, a hexameric ATPase linked to the Mre11 complex in archaeal genomes." EMBO Rep **5**(1): 54-9.

- Martinon, F., A. Mayor and J. Tschopp (2009). "The inflammasomes: guardians of the body." *Annu Rev Immunol* **27**: 229-65.
- Martinon, F. and J. Tschopp (2004). "Inflammatory caspases: linking an intracellular innate immune system to autoinflammatory diseases." *Cell* **117**(5): 561-74.
- Meylan, E., J. Curran, K. Hofmann, D. Moradpour, M. Binder, R. Bartenschlager and J. Tschopp (2005). "Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus." *Nature* **437**(7062): 1167-72.
- Morise, H., O. Shimomura, F. H. Johnson and J. Winant (1974). "Intermolecular energy transfer in the bioluminescent system of *Aequorea*." *Biochemistry* **13**(12): 2656-62.
- Murali, A., X. Li, C. T. Ranjith-Kumar, K. Bhardwaj, A. Holzenburg, P. Li and C. C. Kao (2008). "Structure and function of LGP2, a DEX(D/H) helicase that regulates the innate immunity response." *J Biol Chem* **283**(23): 15825-33.
- Muyldermans, S. (2001). "Single domain camel antibodies: current status." *J Biotechnol* **74**(4): 277-302.
- Muyldermans, S., T. Atarhouch, J. Saldanha, J. A. Barbosa and R. Hamers (1994). "Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains." *Protein Eng* **7**(9): 1129-35.
- Muyldermans, S., C. Cambillau and L. Wyns (2001). "Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains." *Trends Biochem Sci* **26**(4): 230-5.
- Myong, S., S. Cui, P. V. Cornish, A. Kirchhofer, M. U. Gack, J. U. Jung, K. P. Hopfner and T. Ha (2009). "Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA." *Science* **323**(5917): 1070-4.
- O'Neill, L. A. and A. G. Bowie (2007). "The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling." *Nat Rev Immunol* **7**(5): 353-64.
- Ormo, M., A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington (1996). "Crystal structure of the *Aequorea victoria* green fluorescent protein." *Science* **273**(5280): 1392-5.
- Park, H. H., Y. C. Lo, S. C. Lin, L. Wang, J. K. Yang and H. Wu (2007). "The death domain superfamily in intracellular signaling of apoptosis and inflammation." *Annu Rev Immunol* **25**: 561-86.
- Patterson, G. H. and J. Lippincott-Schwartz (2002). "A photoactivatable GFP for selective photolabeling of proteins and cells." *Science* **297**(5588): 1873-7.

- Pichlmair, A., O. Schulz, C. P. Tan, T. I. Naslund, P. Liljestrom, F. Weber and C. Reis e Sousa (2006). "RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates." Science **314**(5801): 997-1001.
- Pippig, D. A., J. C. Hellmuth, S. Cui, A. Kirchhofer, K. Lammens, A. Lammens, A. Schmidt, S. Rothenfusser and K. P. Hopfner (2009). "The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA." Nucleic Acids Res.
- Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast and M. J. Cormier (1992). "Primary structure of the *Aequorea victoria* green-fluorescent protein." Gene **111**(2): 229-33.
- Putnam, C. D., M. Hammel, G. L. Hura and J. A. Tainer (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." Q Rev Biophys **40**(3): 191-285.
- Revels, H., P. De Baetselier and S. Muyldermans (2005). "Nanobodies as novel agents for cancer therapy." Expert Opin Biol Ther **5**(1): 111-24.
- Rhodes, G. (2000). "CRYSTALLOGRAPHY Made Crystal Clear " ACADEMIC PRESS(SECON EDITION).
- Rothbauer, U., K. Zolghadr, S. Muyldermans, A. Schepers, M. C. Cardoso and H. Leonhardt (2008). "A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins." Mol Cell Proteomics **7**(2): 282-9.
- Rothbauer, U., K. Zolghadr, S. Tillib, D. Nowak, L. Schermelleh, A. Gahl, N. Backmann, K. Conrath, S. Muyldermans, M. C. Cardoso and H. Leonhardt (2006). "Targeting and tracing antigens in live cells with fluorescent nanobodies." Nat Methods **3**(11): 887-9.
- Rothenfusser, S., N. Goutagny, G. DiPerna, M. Gong, B. G. Monks, A. Schoenemeyer, M. Yamamoto, S. Akira and K. A. Fitzgerald (2005). "The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I." J Immunol **175**(8): 5260-8.
- Saito, T. and M. Gale, Jr. (2008). "Differential recognition of double-stranded RNA by RIG-I-like receptors in antiviral immunity." J Exp Med **205**(7): 1523-7.
- Saito, T., R. Hirai, Y. M. Loo, D. Owen, C. L. Johnson, S. C. Sinha, S. Akira, T. Fujita and M. Gale, Jr. (2007). "Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2." Proc Natl Acad Sci U S A **104**(2): 582-7.
- Sambrook, J., T. Fritsch and T. Maniatis (1989). "Molecular Cloning A LABORATORY MANUAL." Cold Spring Harbor Laboratory Press.

- Schermelleh, L., P. M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M. C. Cardoso, D. A. Agard, M. G. Gustafsson, H. Leonhardt and J. W. Sedat (2008). "Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy." Science **320**(5881): 1332-6.
- Schmidt, A., T. Schwerd, W. Hamm, J. C. Hellmuth, S. Cui, M. Wenzel, F. S. Hoffmann, M. C. Michallet, R. Besch, K. P. Hopfner, S. Endres and S. Rothenfusser (2009). "5'-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I." Proc Natl Acad Sci U S A.
- Seth, R. B., L. Sun, C. K. Ea and Z. J. Chen (2005). "Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3." Cell **122**(5): 669-82.
- Shaner, N. C., P. A. Steinbach and R. Y. Tsien (2005). "A guide to choosing fluorescent proteins." Nat Methods **2**(12): 905-9.
- Shimomura, O. (1979). "Structure of the chromophore of Aequorea green fluorescent protein." FEBS Letters **104**: 220-22.
- Shimomura, O., F. H. Johnson and Y. Saiga (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea." J Cell Comp Physiol **59**: 223-39.
- Spinelli, S., L. Frenken, D. Bourgeois, L. de Ron, W. Bos, T. Verrips, C. Anguille, C. Cambillau and M. Tegoni (1996). "The crystal structure of a llama heavy chain variable domain." Nat Struct Biol **3**(9): 752-7.
- Spinelli, S., M. Tegoni, L. Frenken, C. van Vliet and C. Cambillau (2001). "Lateral recognition of a dye hapten by a llama VHH domain." J Mol Biol **311**(1): 123-9.
- Sundberg, E. J. and R. A. Mariuzza (2002). "Molecular recognition in antibody-antigen complexes." Adv Protein Chem **61**: 119-60.
- Takahasi, K., M. Yoneyama, T. Nishihori, R. Hirai, H. Kumeta, R. Narita, M. Gale, Jr., F. Inagaki and T. Fujita (2008). "Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses." Mol Cell **29**(4): 428-40.
- Tsai, M. J. and B. W. O'Malley (1994). "Molecular mechanisms of action of steroid/thyroid receptor superfamily members." Annu Rev Biochem **63**: 451-86.
- Tsien, R. Y. (1998). "The green fluorescent protein." Annu Rev Biochem **67**: 509-44.

- Venkataraman, T., M. Valdes, R. Elsby, S. Kakuta, G. Caceres, S. Saijo, Y. Iwakura and G. N. Barber (2007). "Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses." J Immunol **178**(10): 6444-55.
- Wang, S. and T. Hazelrigg (1994). "Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis." Nature **369**(6479): 400-03.
- Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel and D. F. Hochstrasser (1999). "Protein identification and analysis tools in the ExPASy server." Methods Mol Biol **112**: 531-52.
- Xu, L. G., Y. Y. Wang, K. J. Han, L. Y. Li, Z. Zhai and H. B. Shu (2005). "VISA is an adapter protein required for virus-triggered IFN-beta signaling." Mol Cell **19**(6): 727-40.
- Yang, F., L. G. Moss and G. N. Phillips, Jr. (1996). "The molecular structure of green fluorescent protein." Nat Biotechnol **14**(10): 1246-51.
- Yoneyama, M. and T. Fujita (2009). "RNA recognition and signal transduction by RIG-I-like receptors." Immunol Rev **227**(1): 54-65.
- Yoneyama, M., M. Kikuchi, T. Natsukawa, N. Shinobu, T. Imaizumi, M. Miyagishi, K. Taira, S. Akira and T. Fujita (2004). "The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses." Nat Immunol **5**(7): 730-7.
- Zhang, J., R. E. Campbell, A. Y. Ting and R. Y. Tsien (2002). "Creating new fluorescent probes for cell biology." Nat Rev Mol Cell Biol **3**(12): 906-18.
- Zolghadr, K., O. Mortusewicz, U. Rothbauer, R. Kleinhans, H. Goehler, E. E. Wanker, M. C. Cardoso and H. Leonhardt (2008). "A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells." Mol Cell Proteomics **7**(11): 2279-87.

12 Acknowledgments

First of all I would like to thank my supervisor Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work on these most challenging and fulfilling projects. I deeply admire his ingenious way of conducting science. I am very thankful for his constant advice over the years and for leaving me the freedom to find my own way in dealing with scientific problems.

For sure a highlight of my PhD work was the very fruitful and productive collaboration with Dr. Uli Rothbauer and Prof. Dr. Heinrich Leonhardt. It was a very rewarding experience to perform research in a team which combined the tools of structural and cellular biology, resulting in superb scientific findings nobody had expected in the beginning of the project.

I want to express my gratitude to all the members of the Hopfner laboratory. Sharing our every day scientific problems helped a lot in struggling through all this. My special thanks go to Dr. Sheng Cui, who is a great mentor, teacher and friend.

Further acknowledgements go to Prof Dr. Stefan Endres and all the people involved in the DFG-Graduiertenkolleg 1202, who created a very stimulating interdisciplinary research environment and a great resource for scientific exchange.

I am very grateful for the constant support of my family through all the years of studying. A special thanks to all my friends, who made sure that I did not forget that there is a life outside of science.

Last but not least, I deeply thank Natalie for the wonderful PhD time we spent together in Munich. You are the most important part of my life.

13 Curriculum vitae

Personal details

Name	Axel Kirchhofer
Date of Birth	June 26, 1980
Place of Birth	Mönchengladbach, Germany

Education

1990-1999	Gymnasium St. Woflhelm Schwalmtal, Abitur 6/1999
-----------	--

Civil Service

1999-2000	Krankenhausgemeinschaft St. Michael, Schwalmtal
-----------	---

Academic Training

Since 2006	PhD at the Genecenter/Institute of Biochemistry, Ludwig-Maximilians-Universität München in the group of Prof. Dr. Karl-Peter Hopfner
2000-2006	Studies of Biochemistry at the Eberhard-Karls-Universität Tübingen, Diplom 2/2006 with distinction.
2003-2004	Studies of Biochemistry at the California State University, Long Beach, USA