Mapping Innate Immunity pathways: from nucleic-acid recognition to kinase wiring

Doctoral thesis at the Medical University of Vienna
for obtaining the academic degree

Doctor of Philosophy

Submitted by
Adriana Goncalves

Supervisor:
Giulio Superti-Furga, PhD
CeMM, Research Center for Molecular Medicine
Lazarettgasse 14/ AKH BT 25.3 1090 Wien/AUSTRIA

Vienna, February 2012

Acknowledgments

I am truly and deeply grateful to so many people that there is no way to acknowledge them all and no matter how I will express myself it will always seem I am not doing it properly. I would like to express my deepest gratitude to all the ones whose names may not be mentioned here. But I hope that they know that one way or the other they have influenced me in this adventure. A big thank you to you all!

Particularly, I would like to start by expressing my sincere appreciation to my supervisor **Prof. Giulio Superti-Furga**. Giulio has not only been a supervisor but also a mentor to me. He is one of the people who has influenced and inspired me the most in my adult life. I truly believe that I look at, behave and think differently due to his influence. I would like to thank Giulio for giving me the opportunity to work under his supervision. Thank you for all the support, discussions, inspiration, fun moments, understanding and advice.

I would like to express a very special gratitude to **Tilmann Bürckstümmer**. If I would have to mention any other person that has influenced me in so many great ways during the last five years of my life that person would have to be Tilmann. This would not be possible without you. Thank you for always being there when I needed you. This truly means a lot to me. Thank you for caring, thank you for covering my back and thank you for guiding me through these years. I am happy to say that you were not only an advisor but also a friend.

Along with such great experiences come great friends. **Evren Karayel,** my PhD brother, you know that all of this would have been far harder and far less fun if it would not have been for you. We have indeed gone through so many things together and so many of which that only we can understand. Your friendship meant and means a lot to me. Thank you!

I would like to thank, the innate immunity team, the entire GSF lab and all the friends I have made at CeMM for such a creative, intelligent and friendly work atmosphere.

Moreover, I would like to thank my thesis committee members **Sylvia Knapp** and **Caetano Reis e Sousa** for their guidance.

I would like to thank my friends in Portugal for their support and **Francisco Maleno** for standing by me even in the hardest times. You have a special place in my heart.

Finally, I would like to thank my parents and family for being there for me no matter what and for supporting me even when that meant that we would be so many kilometers apart.

Table of Contents

Abstract

Innate Immunity relies on the ability of the molecular machinery of the cell to orchestrate precise molecular interactions for the detection of invading pathogens and the ensuing intracellular and intercellular signaling. Protein kinases, such as TANK-binding kinase 1 (TBK1) and the highly related inducible IκB-kinase (IKK-i), are central components of innate immunity intracellular signaling but are poorly understood in terms of wiring logic. We set out to untangle their molecular interactions as well as those of their known partners TANK, Sintbad and NAP1. We show that the adaptor proteins bind to TBK1 and IKK-i in a mutually exclusive manner suggesting distinct alternative complexes. Binding of each adaptor protein to TBK1 was disrupted by single point mutations on the C-terminal coiled-coil 2 region of TBK1. Using point mutants that affect binding of individual adaptors we found that TBK1 activation in response to double stranded RNA or viral infection was strictly dependent on the integrity of the TBK1/TANK interaction.

TBK1 activation occurs upon recognition of foreign nucleic acids by specific innate immunity receptors. As a complementary approach to identify proteins involved in innate immunity we used immobilized nucleic acids and analyzed binding proteins by mass spectrometry. We identify SAMHD1, a protein associated with Aicardi Goutières Syndrome (AGS), as a nucleic acid binding protein displaying a preference for RNA over DNA. We map a region containing the HD domain as the nucleic acid binding domain. Contrary to wild type SAMHD1, mutants of SAMHD1 observed in AGS patients localize to the cytosol. These data suggest that SAMHD1 has a role in the nucleus that, if disrupted by mutation, leads to cytosolic accumulation of SAMHD1 and autoimmune disease. Overall, the proteomics approach used in my thesis enabled us to identify new players in innate immunity from two different angles of the pathway: TBK1 and IKKi protein complexes; and SAMHD1 as a nucleic-acid interactor. Thus, my thesis provides for an improved understanding of their physical and functional relationship in innate immunity and, potentially, in autoimmune disease.

Zusammenfassung

Das angeborene Immunsystem beruht auf der Fähigkeit der molekularen Maschinerie der Zellen, die zielgerichtete Erkennung eindringender Krankheitserreger auf die darauffolgende intra- und interzellulären Signalweiterleitung abzustimmen.
Obwohl die Proteinkinasen TANK-binding kinase 1 (TBK1) und die ihr eng verwandte, induzierbare IκB-Kinase (IKK-i) zentrale Komponenten der Signaltransduktion der angeborenen Immunität darstellen, ist ihre genaue Position in diesem Signaltransduktionsnetzwerk noch unzureichend erforscht. Aus diesem Grund widmet sich diese Arbeit der Untersuchung der zellulären Bindungspartner von TBK1 und IKKi mit besonderem Fokus auf die Adaptorproteine TANK, Sintbad und NAP1.
Unsere Experimente zeigen, dass die Adaptorproteine an die C-terminale Coiled Coil 2 Region von TBK1 und IKKi binden und dass sich die einzelnen Adaptoren wechselseitig ausschließen. Die Bindung einzelner Adaptoren wurde durch die Einführung einzelner Punktmutationen in der Coiled-Coil 2 Region von TBK1 selektiv unterbunden. Funktionelle Experimente mit den durch Punktmutanten veränderten Adaptorproteinen zeigen, dass die Aktivierung von TBK1 durch doppelsträngige RNA oder Virusinfektion strikt von der Interaktion von TBK1 mit TANK abhängt.
Die Aktivierung von TBK1 erfolgt durch die Erkennung fremder Nukleinsäuren mittels spezifischer Rezeptoren des angeborenen Immunsystems. Zur Identifizierung weiterer, an der angeborenen Immunantwort beteiligter Proteine verwendeten wir immobilisierte Nukleinsäuren und analysierten daran gebunde Proteine mittels Massenspektrometrie.
Wir identifizierten SAMHD1, ein Protein das mit dem Aicardi Goutières Syndrom (AGS) in Verbindung steht. Es handelt sich um ein Nukleinsäure-bindendes Protein, welches präferenziell eher RNA als DNA bindet. Des Weiteren konnten wir die Region des Proteins, welche die HD Domäne enthält, als die Nukleinsäure-bindende Domäne identifizieren. Im Gegensatz zu dem SAMHD1 Wildtyp-Protein befinden sich die in AGS-Patienten gefundenen SAMHD1 Mutanten im Zytosol. Es liegt daher nahe, dass SAMHD1 seine Funktion im Nukleus ausübt, während krankheitsauslösende

Mutationen zur zytosolischen Akkumulation von SAMHD1 und zur Autoimmunerkrankung führen.

Insgesamt hat die in dieser Dissertation verwendete Methode der Proteomik die Identifizierung neuer Akteure des angeborenen Immunsystems ermöglicht, speziell in Hinblick auf die von TBK1 und IKK-i gebildeten Proteinkomplexe weiters wurde SAMHD1 als Nukleinsäuren bindendes Protein identifiziert. Die vorliegende Arbeit ermöglicht auf diese Weise ein besseres Verständnis ihrer physikalischen und funktionellen Beziehungen während der angeborenen Immunantwort sowie ihrer möglichen Rolle im Rahmen einer Autoimmunerkrankung.

1 Introduction

1.1 Immune System

The continuously exposure of organisms to pathogens, demands an efficient protection system. The immune system is the body´s defense mechanism that provides protection against invading microorganisms. It recognizes pathogens such as viruses, bacteria and parasites, as well as other challenges such as tumors and signals derived from damaged cells (immunological recognition) (Matzinger 2002; Medzhitov 2007) and mounts the so called immune response. The immune response involves the coordinated action of specialized cells and molecules that act together to fight and protect against infection (Abbas and Lichtman 2007). Other functions of the immune system include regulation of its own response (immune regulation) and the capability of reacting more rapidly to pathogens that have been encountered previously (Janeway 2008). In higher vertebrates, the immune system is constituted by two major branches: the adaptive immune system and the innate immune system (Pasare and Medzhitov 2004).

1.2 Adaptive Immune System

The adaptive immune system is also known as the specific immune system. As the name indicates, the adaptive immune system recognizes specific protein sequences known as antigens. It is exclusively present in jawed vertebrates and it evolved to be able to react to a great number of antigens. The adaptive immune system has the ability to remember and to respond more promptly in case of recurrent infections - a phenomenon known as immunological memory (Abbas and Lichtman 2007). There are two types of adaptive immune responses: the humoral and the cell-mediated immunity. On one hand, B-lymphocytes produce antibodies and mediate the humoral immunity against extracellular pathogens. On the other hand, intracellular pathogens are dealt by the cell-mediated immunity which is conducted by T-lymphocytes and their secreted cytokines. Both cell types rely on the production of somatically generated receptors to

execute their functions. B-lymphocytes have immunoglobulin receptors at the cell surface which upon activation are secreted as soluble antibodies. Secreted antibodies bind to extracellular microbes and promote their clearance by phagocytosis. (Abbas and Lichtman 2007). Alternatively, intracellular pathogens are fragmented into peptides that are exposed at the cell surface by the major histocompatibility complex (MHC) and recognized by receptors on T-lymphocytes (Janeway 2008).

A pool of naive lymphocytes specific for a large number of antigens can be found even before antigen encounter. During clonal selection, naive lymphocytes recognize antigens displayed by antigen presenting cells (APCs) and get activated into effector lymphocytes.

Effector B-lymphocytes are antibody-secreting plasma cells and effector T-lymphocytes are CD4+ helper T cells or CD8+ cytotoxic T lymphocytes (CTLs) (Abbas and Lichtman 2007). The vital role of the adaptive immune system is illustrated by the problems caused by pathogens that learned how to circumvent the adaptive immune response and genetic disorders leading to immunodeficiency (Janeway 2008).

1.3 Innate Immune System

The innate immune response, also called natural or native immunity, represents the first line of defense to protect the host against disease. The innate immune system employs various effector mechanisms in order to eliminate invading microbes: physical and chemical barriers as the epithelial layer and the respective antimicrobial substances they release, leukocytes (neutrophils, macrophages and natural killer cells), circulating effector proteins (complement, collectins, pentraxins), and cytokines that activate gene expression programs to fight infection (e.g. TNF, IL-1, IL-2, chemokines, type-I Interferons, etc) (Abbas and Lichtman 2007). Most infections are controlled by the innate immune system. Only when the innate immune system is not able to deal with infection, the adaptive immunity is activated. Contrary to adaptive immunity, innate immune recognition is mediated by germline encoded receptors directed against conserved molecular structures of pathogens (Medzhitov 2007).

1.4 Signaling in Innate Immunity

Recognition by the innate immune system occurs through recognition of distinct microbial components by pathogen recognition receptors (PRRs). PRRs are germline-encoded receptors that sense conserved molecular traits of pathogens, the so called pathogen-associated molecular patterns (PAMPs) (Janeway 1989). The functions of PRRs range from opsonization of bacteria and viruses for phagocytosis to triggering of signaling pathways that result in the induction of a variety of antimicrobial genes and cytokines (Medzhitov and Janeway 2000). In addition, the engagement of PRRs leads to the activation of dendritic cells (DCs), allowing the initiation of the adaptive immunity.

There are several functionally distinct classes of PRRs that are known to recognize a wide range of PAMPs in innate immunity. These include the Toll-like receptors (TLRs), the C-type lectin receptors (CLRs), the NOD-like receptors (NLRs); RIG-I-like receptors (RLRs) and cytosolic DNA receptors (Kawai and Akira 2011).

TLRs were the first to be discovered and were named after the protein coded by the Toll gene. The Toll gene is crucial to establish embryonic dorsoventral polarity in *Drosophila,* and was proven to be required for the antifungal response in flies (Lemaitre, Nicolas et al. 1996). TLRs are type-I transmembrane proteins that comprise an ectodomain, a transmembrane region, and a cytosolic TIR (Toll-IL-1 receptor) signaling domain. The ectodomain is constituted by leucin-rich repeats (LRR) and mediates PAMP recognition. On the other hand, downstream signaling is achieved by the TIR domain (Kawai and Akira 2011). So far, 10 different TLR family members have been identified in humans and 12 in mice. Each TLR detects distinct PAMPs: lipoproteins are sensed by TLR1, TLR2, and TLR6; TLR3 recognizes double-stranded (ds) RNA derived from viruses; TLR4, the first mammalian TLR to be identified and characterized (Medzhitov, Preston-Hurlburt et al. 1997; Poltorak, He et al. 1998), recognizes lipopolysacharide (LPS) from gram-negative bacteria; flagellin, the main constituent of bacteria flagella, is recognized by TLR5, single-stranded (ss) RNA is sensed by TLR7 and TLR8, and TLR9 recognizes bacterial DNA containing unmethylated CpG motifs – CpG DNA (Fig. 1.1). Whereas

TLR2 (TLR2/1 or TLR2/TLR6), TLR4 and TLR5 are localized on the plasma membrane, TLR3, TLR7 and TLR9 can be found on endosomes (Fig 1.1).

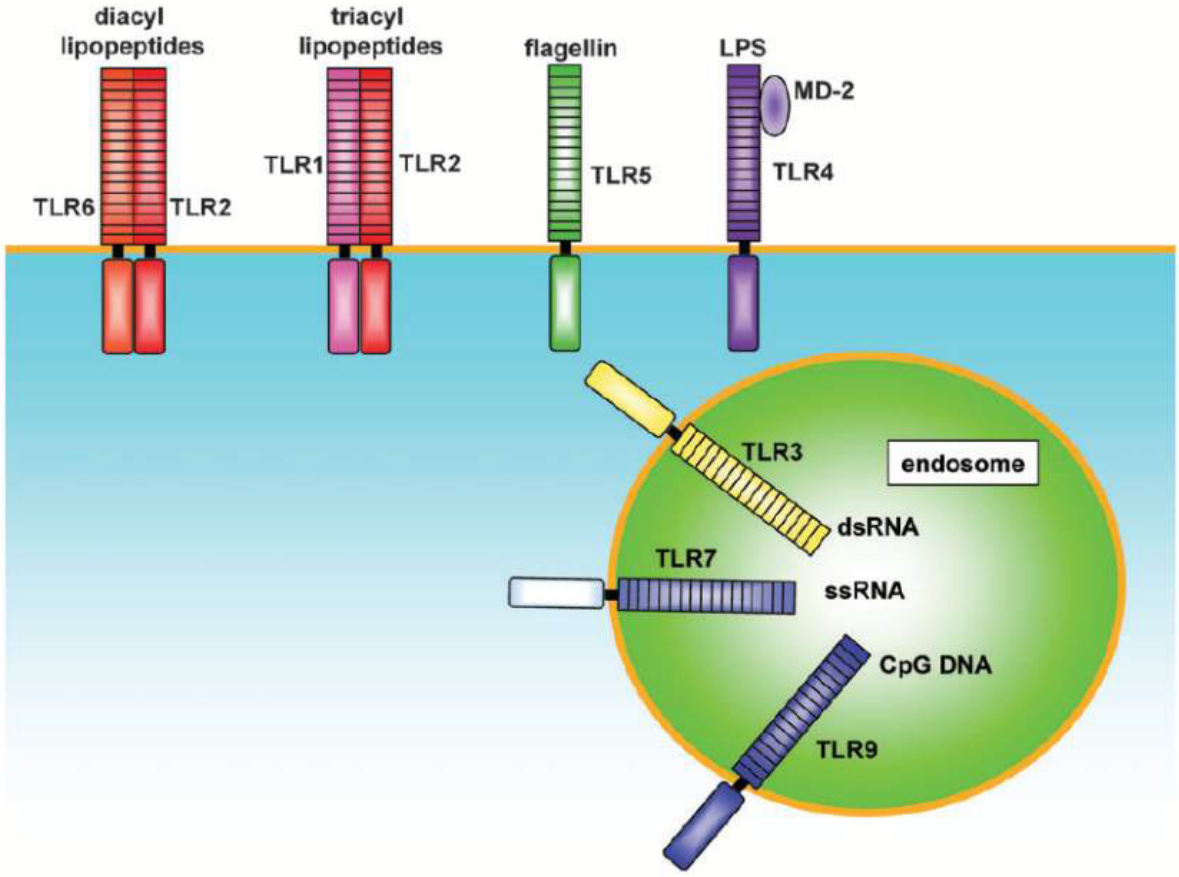

Fig 1.1 Schematic overview of the Toll-like receptors and their ligands. Lipoproteins (recognized by TLR1, TLR2, and TLR6), flagellin (TLR5), LPS (TLR4), dsRNA (TLR3), ssRNA (TLR8) and CpG DNA (TLR9). (Takeda and Akira 2005).

TLR signaling results in the recruitment of adaptor proteins namely MyD88, MAL (TIRAP), TRIF and TRAM to the receptor complex. All the adaptor proteins contain a TIR domain that interacts with the TIR domain of the TLRs leading to activation of distinct signaling pathways. The specificity of TLR signaling is conveyed by the differential recruitment of the different adaptors. MyD88 is used as an adaptor by all TLRs except TLR3 to couple NF-κB and mitogen-activated protein kinases (MAPKs) activation to the production of inflammatory cytokines. On the other hand, TLR3 and TLR4 use TRIF to trigger the production of type-I Interferon and inflammatory cytokines

through activation of the transcription factors NF-κB and IRF3. In general, TLR signaling can be divided into two different pathways depending on which of the adaptors are used: activation of MyD88 leads to the induction of inflammatory cytokines whereas TRIF recruitment drives the stimulation of both interferons as well as inflammatory cytokines. MAL and TRAM work as bridging adaptors to target TRIF to TLR4 and MyD88 to TLR4 and TLR2 (Kawai and Akira 2010), (Fig 1.2).

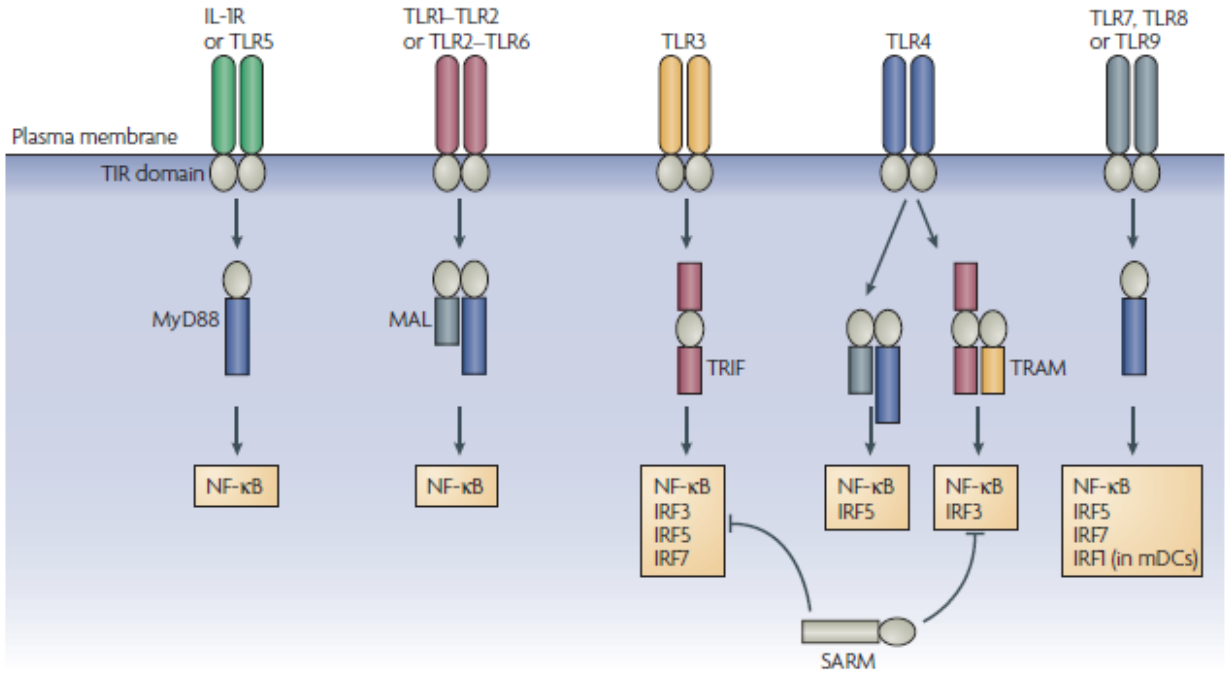

Fig 1.2 Schematic representation of the activation of transcription factors by different TLRs and the differential recruitment of TIR-domain-containing adaptors. MyD88 is required for activation of the IL-1R and all the TLRs except TLR3 that uses TRIF as an adaptor. TLR4 requires both MyD88 and TRIF. MyD88 recruitment leads to NF-κB activation, whereas TRIF activates both NF-κB and IRF transcription factors. MAL and TRAM are co-adaptors for MyD88 and TRIF downstream of TLR1-TLR2 and TLR4 for MAL, and TLR2-TLR6 and TLR4 for TRAM. (O'Neill and Bowie 2007)

Engagement of the MyD88-dependent pathway induces the recruitment of the IL-1 receptors-associated kinases IRAK1, IRAK2, IRAK4 and IRAK-M. IRAK4 is the first kinase to be activated and is followed by the sequential activation of IRAK1 and IRAK2. Activation of these kinases is fundamental for the induction of NF-κB and MAPK

signaling pathways. IRAK activation is followed by the recruitment of the E3 ubiquitin ligase TRAF6. TRAF6 generates K63-linked polyubiquitin chains, which activate the protein kinase TAK1. As a result, NF-κB essential modulator (NEMO) is recruited to the complex leading to a conformational change that allows the activation of the canonical IKKs (IKK-α and IKK-β) by TAK1 and autophosphorylation. Activation of the canonical IKK complex induces NF-κB activation and the production of cytokines. At the same time, TAK1 phosphorylates Erk1, Erk2, p38 and JNK allowing the activation of a number of transcription factors such as AP1. (Kawai and Akira 2010).

Alternatively, stimulation of the TRIF-dependent pathway allows the activation of two other kinases that are closely related to IKK-α and IKK-β, the non-canonical I-κB kinases (TANK)-binding kinase-1 (TBK1) and IKK-i. Upon TBK1 and IKKi activation, the transcription factor IRF3 is phosphorylated, dimerizes and translocates to the nucleus where it triggers the expression of IFN-β and other genes to establish an antiviral program (Clark, Takeuchi et al. 2011).

1.5 Antiviral Immune Responses and type-I Interferon production

Among all the different kinds of pathogens, viruses constitute one of the most abundant and diverse type (Stetson and Medzhitov 2006). Antiviral immunity is dependent on the fast induction of inflammatory cytokines and type-I interferons (IFNs). The production of type-I interferons is mostly regulated at the transcriptional level. Upon viral infection, PRRs recognize viral nucleic acids and activate a signal transduction pathway that ultimately leads to the transcription of the type-I Interferon genes (Pichlmair and Reis 2007). Transcriptional activation of type-I Interferon involves the intricate interplay of several transcription factors, including IRF3/ IRF7, ATF-2/c-Jun and members of the NF-κB family (Panne, Maniatis et al. 2007). Viral infection is controlled by the production of type-I Interferons in a autocrine and paracrine manner. Interferon production activates the type-I Interferon receptor and the induction of interferon-stimulated genes (ISGs) that establish an antiviral state and restrain viral replication.

1.6 TBK1 and IKK-i

TANK-binding kinase 1 (TBK1) and I-κB kinase ε (IKK-ε, also called IKK-i) are pivotal regulators of type-I interferon production: With the exception of plasmacytoid dendritic cells, most cells that are deficient for both TBK1 and IKK-i fail to produce type-I interferons in response to viral infection (Hemmi, Takeuchi et al. 2004; Perry, Chow et al. 2004). Analyses of cells that are single knockouts for TBK1 and IKK-i suggest a certain level of redundancy between TBK1 and IKK-i (Hemmi, Takeuchi et al. 2004), although loss of TBK1 alone seems to have a more profound impact on type-I interferon induction than loss of IKK-i alone (Hemmi, Takeuchi et al. 2004). This is highlighted by the more recent observation that the interferon response to double-stranded DNA depends exclusively on TBK1 and not on IKK-i (Miyahira, Shahangian et al. 2009).

TBK1 and IKK-i are referred to as the "non-canonical I-κB kinases" as they are most closely related to the so called canonical I-κB kinases (IKKs) IKK-α and IKK-β that regulate the activity of transcription factors of the NF-κB family. TBK1 and IKK-i have been shown to activate transcription factors of the IRF family, mainly IRF3, by phosphorylation (Sharma, tenOever et al. 2003). Phosphorylation by TBK1 occurs in the C-terminal domain of IRF3, mainly at serines 386 and 396, and triggers the dimerization and nuclear translocation of IRF3 (Fitzgerald, McWhirter et al. 2003). Alternative substrates of TBK1 include the DEAD-box helicase DDX3X (Schroder, Baran et al. 2008; Soulat, Burckstummer et al. 2008) and phosphorylation of DDX3X is thought to promote IFN-β transcription, but the underlying mechanism is still poorly understood.

TBK1 was identified as a TANK-interacting protein using a two-hybrid screen. TBK1 is a 84 kDa protein that contains an N-terminal serine/threonine kinase domain, a ubiquitin-like domain (ULD) and two C-terminal coiled coil regions (coiled coil 1: residues 619–657, coiled coil 2: residues 682–713).

Similar to TBK1, IKKi is a 80KDa kinase that comprises an N-terminal kinase domain followed by a ubiquitin-like domain (ULD). The C-terminus of IKKi consists of a leucin zipper and a potential helix-loop-helix domain. IKKi was identified as part of an effort to

identify genes that are differentially regulated in response to LPS. IKKi expression occurs mainly in immune cells and is induced upon LPS and proinflammatory cytokines stimulation.

Affinity purification of TBK1 protein complexes led to the copurification of three adaptor proteins of TBK1 named TANK, Sintbad and NAP1 (Bouwmeester, Bauch et al. 2004). TANK was originally identified as a TRAF-binding protein with both stimulatory and inhibitory roles (Cheng and Baltimore 1996). The connection between TANK and TBK1 (TANK-binding kinase 1) only became apparent when TBK1 was found associated with TANK in a yeast-two-hybrid screen for TANK-binding proteins (Pomerantz and Baltimore 1999). Likewise, NAP1 was first isolated in a yeast-two-hybrid screen for NAK1- (or TBK1-) associated proteins (Fujita, Taniguchi et al. 2003). Sintbad, on the contrary, was first found by large-scale proteomics effort (and named TBKBP1) (Bouwmeester, Bauch et al. 2004) and later on characterized in more detail based on its sequence homology to NAP1 (Ryzhakov and Randow 2007). TANK, NAP1 and Sintbad share a common region which mediates association with TBK1 (Ryzhakov and Randow 2007). Loss-of-function experiments using RNAi indicate that all three adaptors are required for production of type-I interferons in response to viral infection (Sasai, Shingai et al. 2006; Guo and Cheng 2007; Ryzhakov and Randow 2007). However, this view was challenged by the recent observation that TANK-deficient mice have no apparent defect in type-I interferon production (Kawagoe, Takeuchi et al. 2009).

1.7 Nucleic Acid Sensing

Contrary to bacteria and Fungi, viruses have no obvious pathogen-specific pattern that is fundamentally different from host components. Evolution has selected for viral receptors that can recognize a trait that is common to all virus – their nucleic acids. Nucleic acid PAMPs can range from virus genomes to replication intermediates. While certain types of nucleic acids, such as dsRNA and 5´triphosphorylated RNA, are recognized based on their nature, others are recognized based on their localization to cellular compartments, such as endosomes and cytosol, where they usually do not occur. Different types of nucleic acid receptors function in distinct cell types and cellular

compartments to mediate the antiviral response (Stetson and Medzhitov 2006; Sharma and Fitzgerald 2011) and the production of type I Interferons (Fig 1.3). There are two complementary receptor systems that detect pathogen nucleic acids: the endosomal Toll-like receptors (TLR3, TLR7, TLR8 and TLR9) and cytosolic receptors which detect viral nucleic acids in endosomes and in the cytosol, respectively.

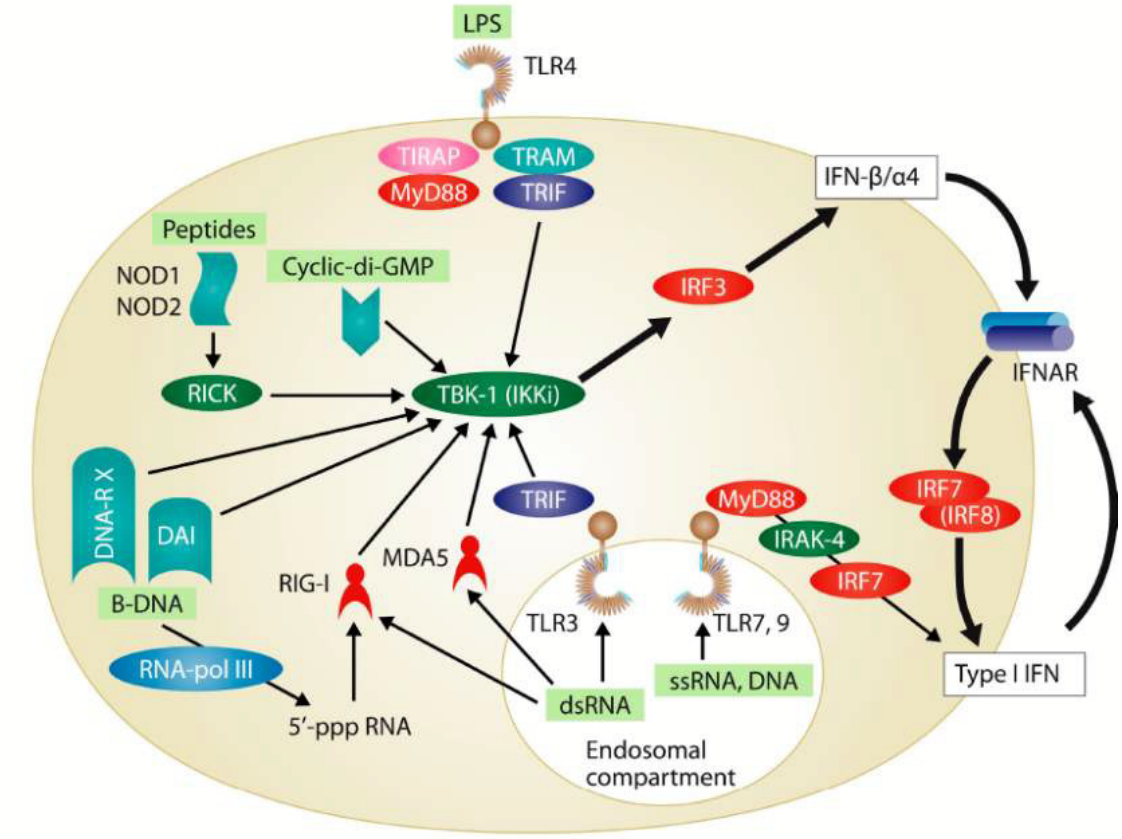

Fig 1.3 Schematic representation of the pathways that lead to type-I Interferon production. In the endosomes TLR7 and TLR9 recognize viral ssRNA (TLR7) and DNA (TLR9) and trigger the production of IFN by engaging the adaptor MyD88. TLR3, also in the endosomes, activates TRIF and mediates IFN production through the TBK1/IRF3 pathway. TRIF is also engaged upon LPS recognition by TLR4 at the plasma membrane. Several intracellular receptors are also known to convey in TBK1 activation and IFN production: the receptors for dsDNA (DAI and others), the RIG-I helicases that recognize 5´ppp RNA, RNA-pol III that converts DNA into 5´ppp RNA that is further recognized by RIG-I, the NLRs and yet an identified receptor of cyclic-di-GMP. (Trinchieri 2010)

The endosomal Toll-like receptors recognize different types of nucleic acids. It is generally believed that, the compartmentalization of nucleic acid-sensing TLRs to

endosomes keeps the host from reacting against self nucleic acid as self nucleic acids do not normally occur in this subcellular compartment. TLRs are largely expressed by macrophages and dendritic cells (DCs). While TLR7 and TLR9 are expressed by both conventional and plasmacytoid dendritic cells (pDCs), TLR3 is mostly expressed by conventional DCs (Stetson and Medzhitov 2006). TLR7 and TLR9 activate a specific signaling pathway that leads to production of high amounts of type I Interferon in plasmacytoid dendritic cells (Colonna, Trinchieri et al. 2004). TLR3 is the receptor for viral dsRNA (Alexopoulou, Holt et al. 2001). TLR3 mediated recognition of dsRNA leads to the production of type I Interferon and proinflammatory cytokines. Single stranded RNA (ssRNA) and virus with ssRNA genomes as for instance vesicular stomatitis virus (VSV), influenza virus and human immunodeficiency virus are sensed by TLR7 and TLR8 in endosomes (Diebold, Kaisho et al. 2004; Heil, Hemmi et al. 2004; Lund, Alexopoulou et al. 2004). TLR7 also senses small antiviral compounds, like imiquimod and loxoribine. TLR9 senses unmethylated DNA CpG motifs that are typical in bacteria and also present in DNA viruses (Hemmi, Takeuchi et al. 2000; Lund, Sato et al. 2003; Krug, Luker et al. 2004).

Although TLRs are expressed in a limited number of cell types, the majority of nucleated cells are able to trigger an antiviral response. This is achieved by activation of germline-encoded cytosolic nucleic acid receptors that trigger the production of type-I Interferons. Cytosolic nucleic acid receptors are expressed ubiquitously and comprise the retinoic acid - inducible gene I (RIG-I)-like receptors (RLRs), and cytosolic dsDNA sensors.

RLRs constituted by RIG-I, MDA5 and LGP2, are a family of DExD/H box RNA helicases that detect RNA from virus in the cytosol and and trigger the production of type-I Interferons and antiviral gene expression to control virus infection (Yoneyama, Kikuchi et al. 2004; Loo and Gale 2011). Even though RLRs are ubiquitously expressed, resting cells produce very low levels of RLRs that are induced after interferon production or viral infection (Kang, Gopalkrishnan et al. 2004; Yoneyama, Kikuchi et al. 2004; Yoneyama, Kikuchi et al. 2005). RIG-I and MDA5 share a similar structure: an N-terminal segment containing tandem caspase activation and recruiting domains (CARD), a DExD/H box RNA helicase domain and a C-terminal regulatory domain

(CTD) which functions as a repressor domain in the case of RIG-I. LGP2 has a similar structure but it lacks the N-terminal CARD domains. RIG-I and MDA-5 detect a large number of viruses. The majority of RNA viruses are recognized by RIG-I (e.g. hepacivirus, members of the Rhabdoviridae, Orthomyxoviridae and Paramyxoviridae), whereas few RNA viruses, such as members of the Picornaviridae, require MDA5. Nevertheless, other viruses namely West Nile Virus and reovirus are detected by both RIG-I and MDA5.

RIG-I and MDA5 employ the CARD domain-containing protein MAVS (IPS-1, VISA or CARDIF) as an adaptor to induce Interferon production. RNA sequences with specific motifs (e.g. dsRNA structure, polyuridine motifs with interspersed C nucleotides) that are marked with 5´triphosphorylated (5´ppp) ends are recognized by RIG-I (Hornung, Ellegast et al. 2006; Pichlmair, Schulz et al. 2006; Loo and Gale 2011). It is believed that the 5'triphosphorylated open end of viral RNA is the key factor that allows discrimination between self and non-self RNA. In mammal RNAs, the 5'ppp end is not available for recognition by RIG-I: it is either obscured by ribosomal proteins as in the case of rRNA, or removed or modified from mRNAs and tRNAs before they reach the cytosol (Ramakrishnan 2002; Pichlmair and Reis e Sousa 2007). Ligand binding to RIG-I induces a conformational change that activates the formation of a protein complex and mediates the induction of the type-I Interferon response.

Even though the fact that DNA is a potent inducer of inflammatory responses is known for over a decade, mechanisms underlying the sensing of DNA remain poorly understood. The recognition of dsDNA in the cytosol, leads to the cleavage of pro-IL-1β into active IL-1β and to the production of IFN-β and other cytokines by a number of intracellular receptors (Ishii, Coban et al. 2006; Stetson and Medzhitov 2006; Pichlmair and Reis e Sousa 2007; Muruve, Petrilli et al. 2008).

The cytosolic DNA recognition system can be divided in three pathways: (1) DNA can be recognized by cytosolic DNA receptors to activate the production of type I Interferons and proinflammatory cytokines, (2) DNA can be converted to RNA by RNA polymerase III and activate RIG-I to trigger a similar transcriptional program, and (3) DNA can be

recognized by AIM2 to form an active inflammasome that leads to the processing of pro-IL-1β and pro-IL-18 to mature IL-1β and IL-18 release (Hornung and Latz 2010).

Receptors involved in cytosolic DNA recognition include: DNA-dependent activator of interferon (IFN)-regulatory factors (DAI) (also called Z-DNAbinding protein 1, ZBP1) (Takaoka, Wang et al. 2007), absent in melanoma 2 (AIM2) (Burckstummer, Baumann et al. 2009; Fernandes-Alnemri, Yu et al. 2009; Hornung, Ablasser et al. 2009; Roberts, Idris et al. 2009), RNA polymerase III (Pol III) (Ablasser, Bauernfeind et al. 2009; Chiu, Macmillan et al. 2009), leucine-rich repeat (in Flightless I) interacting protein-1 (Lrrfip1) (Yang, An et al. 2010), DExD/H box helicases (DHX9 and DHX36) (Kim, Pazhoor et al. 2010), and the IFN inducible protein IFI16 (Unterholzner, Keating et al. 2010).

Recently NLRP-3, a member of the NODlike receptor (NLR) family, was shown to regulate the production of active IL-1β in response to the DNA viruses, adenovirus and Herpes simplex virus (HSV) (Muruve, Petrilli et al. 2008; Davis, Wen et al. 2011). Nevertheless, NLRP-3 is not required to mediate the response to double stranded cytosolic DNA. The NLR family members share a similar domain structure that comprises a nucleotide-binding domain (NBD) and a leucine-rich repeat (LRR) domain that is also present in other innate immunity receptors as for instance the TLRs. The caspase activation and recruitment domain (CARD) and the pyrin domains (PYDs) are other characteristic domains that can be found in the NLR proteins. Through homotypic interactions, these domains recruit and activate adaptor proteins and downstream signaling effectors. NLRs are activated by a wide range of pathogen- and danger-associated molecular patterns (PAMPs and DAMPs respectively). Many NLRs, such as NLRP1 (NALP1) and NLRP3 (NALP3), associate with the adaptor ASC leading to the formation of an active inflammasome that activates Caspase-1 to further cleave pro-IL-1β into mature IL-1β. DNA stimulation or vaccinia virus infection leads to activation of caspase-1 and successive secretion of IL-1β in an ASC-dependent manner. Once NLRP1, NLRP3, NLRP6 and NLRP12 were excluded as putative DNA receptors, several independent groups (Burckstummer, Baumann et al. 2009; Fernandes-Alnemri, Yu et al. 2009; Hornung, Ablasser et al. 2009) identified AIM2 as a DNA sensor in the inflammasome pathway.

DAI/ZBP1 was the first protein to be proposed as an IFN inducing cytosolic DNA receptor (Takaoka, Wang et al. 2007). DAI is an IFN inducible protein itself which is able to bind B-DNA and whose expression increases the production of pro-inflammatory cytokines and type I IFNs in response to B-DNA. Nevertheless, as DAI deficient mice have a normal interferon response when challenged with dsDNA or DNA virus (Ishii, Kawagoe et al. 2008), the role of DAI in DNA sensing remains controversial. At the same time, even the RNAi phenotype seems to be restricted to certain cell types: While DAI is necessary for dsDNA mediated IFN induction in L929 fibroblasts, it is dispensable in mouse embryonic fibroblasts (Wang, Choi et al. 2008). These results suggest that DAI has only a redundant and cell-type specific role in DNA recognition.

More recently, two independent studies have demonstrated that AT-rich DNA in the cytosol triggers the production of type I Interferons through RNA Polymerase III that transcribes AT-rich DNA into RNA that can be further recognized by RIG-I (Ablasser, Bauernfeind et al. 2009; Chiu, Macmillan et al. 2009). However, the physiological relevance of this finding is currently unclear as this response is largely dependent on the use of the artificial DNA ligand poly(dAdT).

In plasmacytoid dendritic cells, DHX9 and DHX36 are required for cytosolic CpG-DNA and HSV-1 mediated interferon response (Kim, Pazhoor et al. 2010).

LRRFIP1 has been identified to bind both RNA and DNA. However, contrary to other receptors, LRRFIP1 exerts its functions by recruiting β-catenin to enhance IFN-β response (Yang, An et al. 2010). Upon association with LRRFIP1, β-catenin is phosphorylated and translocates to the nucleus where it leads to the recruitment of histone acyltransferase p300 to the IFN promoter and further transcriptional activation of IFN-β. The LRRFIP1/β-catenin pathway is triggered by DNA from *Listeria monocytogenes,* an intracellular bacterial pathogen, and RNA from vesicular stomatitis virus (VSV), a single-stranded RNA virus.

Another potential DNA receptor is IFI16 which binds to a DNA motif derived from vaccinia virus and leads to the induction of IFN-β and inflammatory cytokine production (Unterholzner, Keating et al. 2010). Similar to AIM2, IFI16 is a PYHIN protein that uses

the HIN domain to bind DNA; however, unlike AIM2, IFI16 is a predominantly nuclear protein and for this reason had been discarded as potential cytosolic DNA sensors in the past (Burckstummer, Baumann et al. 2009). More recently, Unterholzner and colleagues have shown that a pool of IFI16 resides in the cytoplasm where it recognizes viral dsDNA leading to type-I Interferon production. Loss-of-function experiments demonstrated that dsDNA-mediated interferon induction requires IFI16, TBK1 and STING. Furthermore, IFI16 activation leads to the production of IFN-β and proinflammatory cytokines in response to cytosolic viral DNA stimulation or HSV1 infection (Unterholzner, Keating et al. 2010). In contrast, Kerur and colleagues demonstrated that during nuclear replication of Kaposi sarcoma-associated herpesvirus (KSHV), IFI16 associates with ASC and procaspase-1 to form an active inflammasome that activates the production of IL-1β (Kerur, Veettil et al. 2011). The authors suggest that IFI16 recognizes KSHV in the nucleus and further relocalizes to the cytosol where it forms a functional inflammasome. These results suggest that IFI16 may be able to trigger both type-I Interferons as well as IL-1β production depending on the cellular compartment where recognition occurs.

1.8 Autoimmune Disease

The immune system has to be able to attack and eliminate a wide range of pathogens which themselves can adapt to evade the immune system of their hosts. In order to achieve this, the immune system has evolved mechanisms to generate a vast immunological repertoire to ensure that all microbes are attacked one way or the other.

It is however very important that the immune system discriminates between self and non-self antigens. The collection of receptors involved in immune recognition, has to be able to distinguish between pathogens and the molecules that constitute our own bodies. Such type of complex system needs tight regulation and breakage of self tolerance is the fundamental cause of autoimmunity. Self tolerance can be classified as either central tolerance or peripheral tolerance.

Central tolerance occurs by eradication of self-reactive immature lymphocytes in primary lymphoid organs. Immature B- and T-cells that recognize self antigens are eliminated by apoptotic cell death in the bone marrow and the thymus respectively. Peripheral tolerance takes place upon recognition of self antigens by mature lymphocytes in peripheral lymphoid organs. Mature lymphocytes that react against self antigens are either killed or deactivated. Peripheral tolerance is achieved through anergy (lack of reaction), deletion (apoptosis) or suppression by regulatory T cells. Breakage of tolerance and activation of self-reactive lymphocytes leads to the development of autoimmune disease. Autoimmune disease is a multigenic disorder that is believed to result from the interaction of genetic, environmental and regulatory factors (Rioux and Abbas 2005).

Autoimmune disease can be caused by a deregulation of nucleic recognition that is either genetic or determined by the environment. Under normal conditions, self-derived nucleic acids do not induce an innate immune response. Nevertheless, contrary to other PAMPs that are microbe specific, nucleic acids are an integrant part of the host. How does the immune system avoid recognition of self nucleic acids? On the one hand, specific localization of innate immune receptors, as in the case of endosomal TLRs, avoids recognition of self ligands by the immune system. On the other hand, self nucleic acids are often degraded by nucleases avoiding their recognition by innate immune receptors. However, malfunctions of the immune system can lead to recognition of self nucleic-acids and consequent unwanted toxicity, pathological damage, inflammation and autoimmune disease.

1.8.1 Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) was among the first autoimmune disorders to be associated with abnormal nucleic acid recognition. SLE is characterized by microvascular inflammation, tissue damage, antibodies against self nucleic acids or nucleoproteins, and high levels of type I IFN. The elevated levels of type I Interferons in the serum of patients with SLE correlate with disease. Moreover, pDCs produce type-I Interferon in response to serum of SLE patients (Kawai and Akira 2010).

Immune complexes with self-derived nucleic acids, nucleoproteins and autoantibodies are internalized by Fc receptors on dendritic cells. Internalization of such complexes into vesicles containing TLR-7 and TLR9 trigger the activation of TLR signaling and the production of type I interferon (Means, Latz et al. 2005; Vollmer, Tluk et al. 2005). Under healthy conditions, those nucleic acids are removed by nucleases present in different subcellular compartments. Mutations in some of those nucleases have been linked to the development of SLE (Fig 1.4).

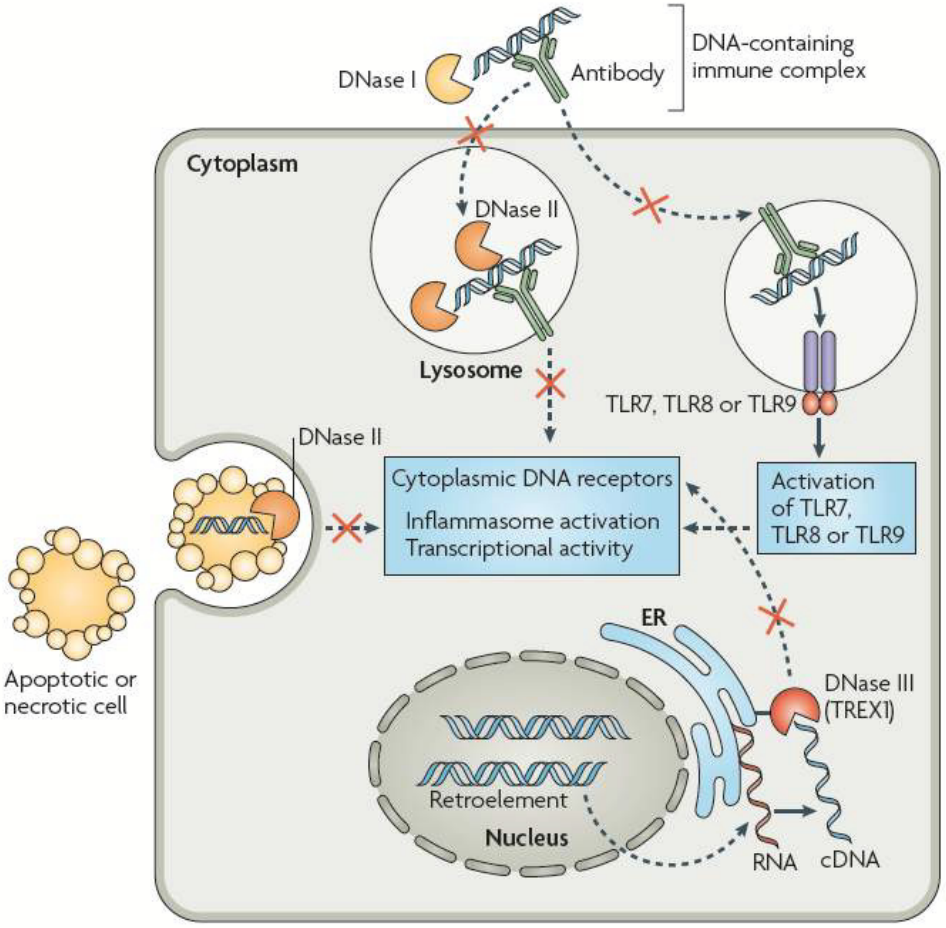

Fig 1.4 Mutations in different nucleases lead to accumulation of endogenous nucleic acids and autoimmune disease (Hornung and Latz 2010)

DNA localization is normally restrained to the nucleus and mitochondria. DNA that is found in the cytosol or other cellular compartments is removed by digestion by different

DNases. Extracellular DNA that arises from apoptotic cells can be digested by DNase I which is normally present in the exracellular space. Mutations in the gene that encodes DNase I are associated with SLE in humans and cause a similar pathology in mice (Napirei, Karsunky et al. 2000; Yasutomo, Horiuchi et al. 2001; Barber 2011). In the lysosomes, the engulfment of apoptotic and necrotic cells leads to DNA release that is cleared by DNase II. DNase II deficient mice succumb from elevated levels of type I Interferons (Yoshida, Okabe et al. 2005). TREX1 has been recently described as a SLE associated exonuclease that digests DNA reverse transcribed from retroelements, avoiding their accumulation in the cytoplasm (Stetson, Ko et al. 2008).

1.8.2 Aicardi-Goutières Syndrome

Aicardi-Goutières syndrome (AGS) is a rare genetic autoimmune encephalopathy whose features phenocopy congenital viral infection. It consists of a familial progressive encephalopathy associated with calcifications of the basal ganglia, white matter abnormalities, a chronic cerebrospinal fluid (CSF) lymphocytosis, and elevated levels of IFN-α in CSF and negative serological evidence for common prenatal infections. In line with this, AGS is characterized by elevated type-I Interferon production (Crow 2007; Crow and Livingston 2008; Rice, Bond et al. 2009). The incidence of AGS is very low. Onset occurs before 3-7 months of age in most patients. Mutations in genes encoding the exonuclease TREX1 and the three subunits of the RNase H2 endonuclease complex (RNase H2A, RNase H2B, RNase H2C) have been reported to cause AGS (Crow, Hayward et al. 2006; Crow, Leitch et al. 2006; Rice, Newman et al. 2007; Rice, Bond et al. 2009). At least in the case of TREX1, data suggest that protein dysfunction leads to the accumulation of endogenous nucleic acids, which are recognized by innate immune receptors and trigger the production of type-I Interferon (Crow, Leitch et al. 2006; Yang, Lindahl et al. 2007; Stetson, Ko et al. 2008).

More recently, SAM domain and HD domain containing protein 1 (SAMHD1) has also been associated with AGS (Rice, Bond et al. 2009). SAMHD1 comprises a sterile alpha motif (SAM) and an HD domain. SAM domains are putative protein-protein interaction units present in a broad variety of proteins that can either homo- or hetero-oligomerize

(Kim and Bowie 2003; Qiao and Bowie 2005). HD domains, named after a conserved doublet of histidine (H) and aspartate (D), are found in a diverse family of phosphohydrolases. AGS-causing SAMHD1 mutations have been observed to occur throughout the entire gene, resulting in single amino acid substitutions as well as truncations and deletions (Rice, Bond et al. 2009). Plausibly, SAMHD1 may act as an unconventional nuclease by analogy to TREX1 and RNASEH2, but nuclease activity has not been described.

SAMHD1 was originally identified as an ortholog of the IFN-γ-induced murine gene *Mg11* and was referred to as dendritic cell–derived IFN-γ–induced protein (DCIP) (Li, Zhang et al. 2000). It is transcriptionally upregulated in response to viral infection in an interferon-dependent manner, and has been implicated in proinflammatory responses (Li, Zhang et al. 2000; Prehaud, Megret et al. 2005; Hartman, Kiang et al. 2007; Liao, Bao et al. 2008). Very recently, SAMHD1 has been shown to be an HIV-1 restriction factor and the cellular target of the HIV-2 protein Vpx (Laguette, Sobhian et al. ; Hrecka, Hao et al. 2011). Nevertheless, the mechanism by which dysfunction of SAMHD1 results in autoimmune disease is still poorly understood.

2 Aim of the Study

In most cells, antiviral immunity relies on the recognition of viral nucleic acids by different intracellular receptors that trigger the activation of the kinases TBK1 and IKK-i and the production of type-I Interferon. We aimed at understanding the antiviral innate immunity pathways starting from the first step of viral recognition to the wiring of downstream kinases that ultimately lead to the production of type-I interferons and the establishment of an antiviral state. More specifically, we undertook a systematic proteomics approach to identify new players in innate immunity from two different angles of the pathway: a) TBK1, IKK-i and their respective adaptor proteins TANK, Sintbad and NAP1, and b) proteins involved in the recognition of viral nucleic acids.

Regarding the kinases TBK1 and IKK-i, the aims of the study involved: i) identification of the protein complexes around TBK1, IKK-i and their adaptor proteins, ii) understanding the structural requirements for TBK1 activity iii) understanding the physical and functional relationship of the TBK1 adaptors with TBK1 and with one another.

Next, we aimed at identifying proteins involved in the recognition of viral nucleic acids and understanding their role in innate immunity. We identify nucleic acid binding proteins by affinity-based proteomics. As selection criteria for validation, we prioritize proteins that share common features to known players in innate immune recognition such as inducibility by innate immunity stimuli and/or involvement in autoimmune disease.