**Evolution of antiviral breadth in Mx GTPases**

Patrick S. Mitchell

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Harmit S. Malik, Chair

Michael Emerman

Daniel B. Stetson

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

## Abstract

Evolution of antiviral breadth in Mx GTPases

Patrick S. Mitchell

Chair of the Supervisory Committee:
Assistant Member Harmit S. Malik
Department of Molecular and Cellular Biology
Fred Hutchinson Cancer Research Center
Howard Hughes Medical Institute

As obligate parasites, viral pathogens coopt host resources and subvert cellular processes to promote their own replication. In response, host organisms have evolved numerous countermeasures to thwart viral infection. However, the molecular basis for how host antiviral defenses overcome the daunting challenge of viral diversity and the highly adaptive nature of viral pathogens is poorly understood. Viral pathogens and their infected hosts are engaged in a constant battle to gain evolutionary dominance. Such host-virus “arms races” drive the rapid evolution of genes in conflict to gain a fitness advantage through recurrent innovation. Herein, I leverage evolutionary signatures of these adaptive processes to gain insight into molecular mechanisms by which the broad-acting antiviral protein MxA overcomes the challenge of viral pathogen diversity. This evolution-guided approach identified and allowed the characterization of multiple surfaces on MxA that function as independent modules to define target recognition and antiviral specificity. These studies provide an evolutionary and molecular basis for MxA

antiviral breadth, and suggest general principles by which cell-intrinsic immunity can tip the balance against rapidly evolving RNA viruses.

# Table of Contents

**List of Figures**

**List of Tables**

**Acknowledgement**

I acknowledge and am grateful for support from the NSF GRFP, the NIH Interdisciplinary Training Grant (T32 CA 080416), the Howard Hughes Medical Institute and the University of Washington Cellular and Molecular Biology Program, from the American Society for Virology for travel awards and the American Society for Microbiology for the honor of receiving the Raymond W. Sarber award. I am very fortunate to have had incredible administrative support from Michele Karantsavelos, MaryEllin Robinson, Carolyn Goard, Alex Moreno and Maia Low. I want to thank Dan Stetson, Jason Smith and Evan Eichler for their guidance, expertise and service on my thesis committee. I thank Otto Haller, Georg Kochs and Corinna Patzina for a wonderful collaboration. I am also grateful for local collaborators, especially Adam Geballe, who provided numerous insights and resources throughout my training.

I have been especially fortunate during my graduate training to be surrounded by an amazing group of people. I am deeply appreciative for the greater Hutch community – it is impossible to imagine a better scientific culture to have been raised in. In particular, I want to thank Katie Peichel, Dan Gottschling, Adam Geballe, Sue Biggins, Beverly Torok-Storb and of course Michael and Harmit for exemplifying the special ethos and character of the Center.

Harmit and Michael are wonderful scientists and amazing people. Words fail to capture how truly thankful and profoundly fortunate I am to have had them as mentors. I am particularly appreciative of their enduring patience in allowing me the freedom to explore various avenues of research while in their labs, but I am most grateful for the training environment that they have provided. To all my friends and colleagues in the Malik and Emerman Labs, thank you for everything.

Finally, to my mom, dad and sister, to my dear friends, and most of all to Alana: I am deeply grateful for you, and fully indebted to you for making my life rich and meaningful. Thank you.

## Chapter 1: Introduction

Parts of this Chapter were modified from (*1*) with permission, Elsevier license number 3591720358649.

### *General Introduction*

Like the remarkable diversity in primate morphology, the antiviral repertoire of closely related species is also rapidly evolving. This marked divergence in host determinants of cell-intrinsic immunity is an indelible imprint from a legacy of lineage-specific arms races between pathogenic viruses and their infected hosts. The resultant inter- and intra-species genetic variation underlies differences in the molecular barriers to viral infection, and ultimately influences viral host range and the related phenomena of viral emergence. Evolutionary signatures of host-virus conflicts mark key innovations that have allowed host genomes to successfully overcome past pandemics. Therefore retrospective analyses of these adaptive processes from our evolutionary history can reveal solutions to our current susceptibility to extant pathogens.

*The cell-intrinsic response to viral infection*

Viruses are the most abundant genetic entity on Earth. As obligate parasites, viruses depend on host resources and coopt cellular processes to drive their own replication. In response, an astonishing diversity of mechanisms to thwart viral infection has evolved throughout the tree of life.

Cell autonomous or intrinsic immunity is an ancient and pervasive form of host defense against viral pathogens. In both prokaryotes and eukaryotes, the ability to distinguish between self and non-self is critical for the deployment of an appropriate response, a concept brought forth by Charles Janeway nearly 25 years ago (*2*). For example, in bacteria and Achaea, integrated fragments of foreign nucleic acids (i.e., CRISPR, clustered regularly interspaced short palindromic repeats) are transcribed to generate small RNA guides (crRNAs) that target large nuclease-containing antiviral complexes (i.e., Cas protiens) to invading cognate bacteriophage or plasmid (*3*-*5*). Similarly, plants and invertebrates utilize virus-derived dsRNA to generate Dicer substrates, which in turn directs RNA-mediated interference (RNAi) against viral genomes or messenger RNAs (*6*-*8*). Although functionally analogous, effector proteins of the CRISPR/Cas and RNAi are non-orthologous, highlighting the remarkable complexity of host defense systems even when converging upon common themes.

Small RNAs play an essential role in bacterial and invertebrate detection and response to pathogens, however, protein-based sensors and effectors (e.g., restriction endonucleases) are also common. These represent the predominant form of pathogen detection in vertebrate cell-intrinsic immunity. Although the evolutionary pressures that ultimately selected for a shift from RNA-centric to protein-centric immunity is unknown, a commonality between these systems is the recognition of foreign nucleic acids. In vertebrates, viral RNA and DNA represent the most

abundant ligand in a class of pathogen-associated molecular patterns (PAMPs) that are recognized by host pattern recognition receptors (PRRs) to distinguish self from non-self.

Classical PRRs that detect viral nucleic acids can be categorized based on cellular compartmentalization: the endosomal toll-like receptors (TLRs), and the cytosolic RIG-I-like receptors (RLRs), nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and AIM2-like receptors (ALRs) (reviewed in (*9, 10*). Recently, PRRs such as gamma-interferon-inducible protein 16 (IFI16) (*11*) and Mab-21 domain containing (Mb21d1, or cGAS) (*12-14*) have been identified, which also bind pathogen-derived nucleic acids. Together, this array of sensors act as sentinels for detection of RNA and DNA viral genomes or nucleic acid products of replication. A major outcome of ligand binding by PRRs is the induction of cytokines. Formation of NLR and ALR inflammasomes promote caspase-1 activation and subsequent cleavage of pro-inflammatory cytokines interleukin (IL)-1B and IL-18, thereby linking the intrinsic and innate responses to viral infection. Cell-intrinsic immunity against RNA viruses is largely mediated through TLR and RLR signaling. Although each class of sensors act through distinct signaling cascades, both pathways ultimately converge on the transcription factors nuclear factor kappa B (NFkB) and interferon regulatory factor 7 (IRF7) or IRF3, respectively, leading to the transcriptional activation of type I interferon (IFN). Type I IFN acts in both an autocrine and paracrine fashion through interaction with its cognate receptor (i.e., IFNAR, interferon-alpha/beta receptor). IFN signals through JAK/STAT intermediates to promote the translocation of IRF9, which together target IFN-stimulated response elements (ISRE) to induce interferon-stimulated genes (ISGs).

ISG products are the effectors of the cell-intrinsic response to viral infection, acting in

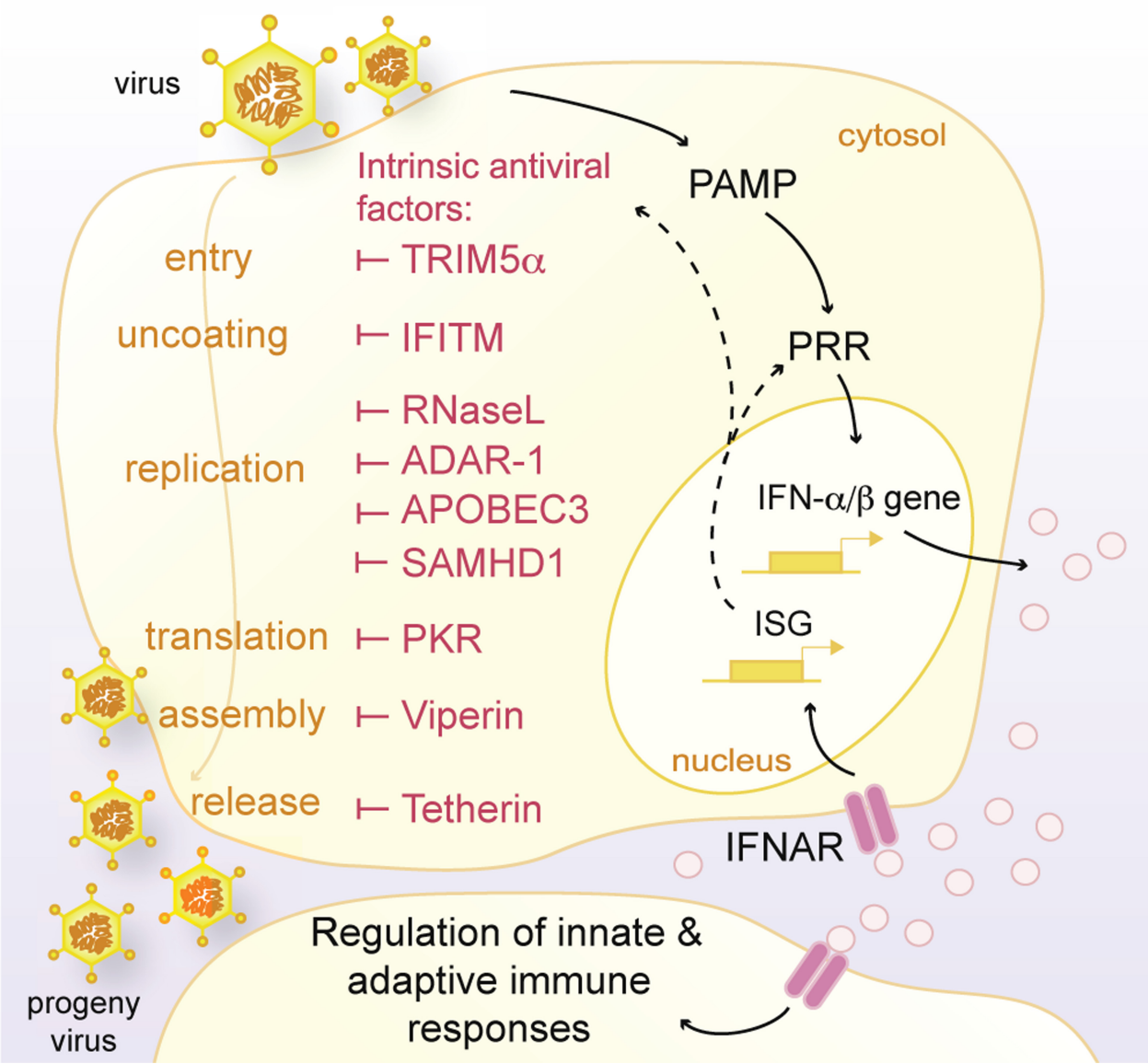



**Figure 1. The cell-intrinsic response to viral infection.**

A schematic representation of the IFN-mediated "antiviral state" in a virus-infected cell. Cell-intrinsic immunity is established upon binding of pathogen-associated molecular patterns (PAMPs) by host pattern recognition receptors (PRRs). Substrate binding by PRRs activates various signaling cascades that ultimately result in the induction of type I IFNs and the transcriptional activation of hundreds of interferon stimulated genes (ISGs). ISGs act as molecular barriers to viral infection, blocking various aspects of viral replication. **Figure 1** was reproduced from (*15*) with permission, Elsevier license number 3591701135464.

concert to establish the "antiviral state" (**Figure 1**). This germline-encoded defense strongly influences host permissiveness to viral infection. For example, IFNAR$^{-/-}$ mice are profoundly more permissive to viral infection relative to wild type littermates (*16*). The genetic origin of

some human immunodeficiency syndromes have also been linked to mutations in key modulators of the IFN-response (*17, 18*). The fitness cost that PRRs and ISGs impose on viruses is highlighted by the myriad means by which viruses antagonize or subvert these host responses (*19*). For example, the highly pathogenic avian influenza virus HPAI and coronavirus SARS-CoV induce repressive histone modifications to dampen the expression of ISGs (*20*). Interestingly, correlations between uncoating defects and enhanced sensitivity of lentiviruses to interferon suggests that a general function of ISGs may be to alter replication kinetics and thereby provide a greater opportunity for detection by PRRs. Importantly, the inappropriate or hyper activation of PRRs and ISGs are an important genetic component of human autoimmune-related diseases, or "interferonopathies" (*21*), and contribute significantly to pathogenesis associated with viral infection.

*PRR-like substrate preference of ISGs is the molecular basis for antiviral specificity*

Recent strategies to systematically characterize the antiviral function of the ISG repertoire have revealed effectors that act in either a broad or targeted fashion (*14*). This characteristic difference in antiviral breadth is a product of the type of substrate engaged by ISGs to enact effector function. There are several, well-studied examples of broad recognition and response to viral infections. This may reflect that although the "antiviral state" includes hundreds of ISGs, there still exists an imbalance when considering the diversity of viruses and the significantly higher mutation rates that allow for rapid adaptation to host defenses. Broad antiviral activity of some ISGs may partially compensate for this inequity. For instance, Protein Kinase R (PKR) detects double-stranded cytoplasmic RNA, a product of many viruses, which

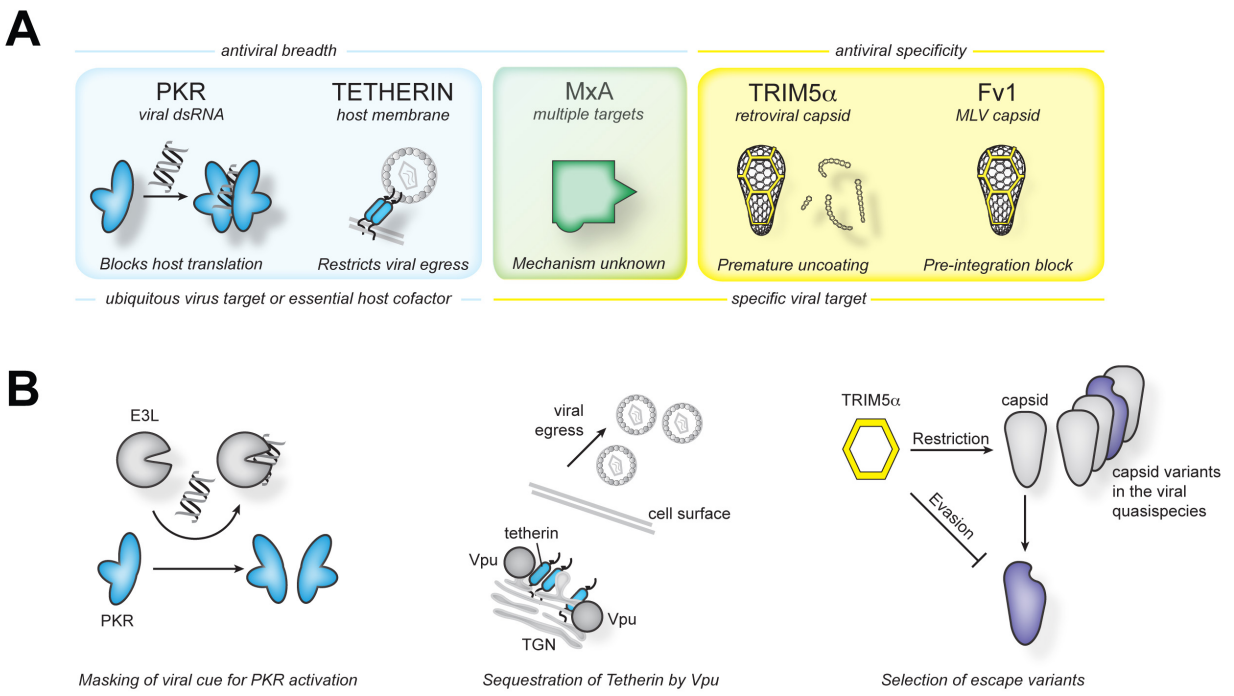



**Figure 2. Target recognition determines the range in antiviral activity**

**A.** Depicted are illustrative examples of broad (shown in blue) and narrow (shown in yellow) acting antiviral proteins. Broad acting antiviral factors tend to act through recognition of ubiquitous viral substrates. For example, double-stranded RNA sensing activates Protein kinase R (PKR), which blocks host protein synthesis through the phosphorylation of eIF2alpha (*22*). Broad-acting antiviral factors can also act by targeting cellular substrates that are essential for viral replication. For example, tetherin incorporates into host membranes, effectively 'tethering' budding virions to the cell surface (*23*). In contrast, the recognition of specific substrates narrows antiviral specificity. TRIM5alpha binds the retroviral capsid lattice, which promotes premature uncoating (*24*). In turn, TRIM5alpha shows strong specificity for simian retroviruses. Similarly, murine Fv1 restricts B-tropic murine leukemia viruses (B-MLV) with exquisite specificity, which limits its activity against even the highly related N-tropic MLV (*25*) (*26*). MxA is shown as a blend of these classification schemes (green), achieving antiviral breadth through recognition of distinct viral targets that vary across viral families. **B.** Common viral evasion strategies in response to broad or narrow acting antiviral factors. Viruses often encode antagonists that either directly or indirectly overcome broad-acting antiviral factors. For example, the poxvirus-encoded antagonist E3L masks viral dsRNA thereby preventing PKR activation (left). Alternatively, HIV-1 Vpu sequesters tetherin at the transgolgi network (TGN) preventing its trafficking to the plasma membrane (middle) (reviewed in (*27, 28*)). Host factors with specific targets can be overcome by selection of viral variants that carry escape mutations (right).

blocks protein production via phosphorylation of the eukaryotic initiation factor 2a (eIF2α) (**Figure 2A**) (*22*). Similarly, ISGs can leverage the obligate nature of viruses by targeting cellular components essential to viral replication. For example, the host restriction factor BST2/TETHERIN inhibits a wide range of enveloped viruses by sequestering budding virions to the cell through non-specific interactions with host and viral membranes (**Figure 2A**) (*23*). The strategy of targeting essential and ubiquitous features of viral replication is likely effective because it precludes the selection for adaptive "escape" mutations by establishing a prohibitively high fitness cost. Viral accessory proteins that function as direct antagonists of ISGs best exemplify the effectiveness of this strategy. For example, the poxvirus vaccinia virus encodes two PKR antagonists that either masks the PKR substrate by sequestering dsRNA (*29*), or structurally mimics and thereby directly competes for PKR binding with eIF2α (*30*). Similarly, lentiviruses like HIV-1 encode the accessory protein Vpu, which degrades and/or sequesters TETHERIN (**Figure 2B**) (*31, 32*).

In contrast to these generalist strategies, some antiviral factors show strong pathogen-specificity as a result of targeting specific viral components (**Figure 2A**, yellow). For instance, Tripartite motif-containing protein 5a (TRIM5α) specifically binds to incoming retroviral capsid assemblies, thereby promoting premature viral uncoating (*24*). The Friend virus susceptibility 1 (Fv1) antiviral factor is another example of a highly specific capsid-interacting protein whose antiviral range and potency is determined entirely by its binding specificities (*25*). These factors provide robust protection against incoming viruses. Indeed, TRIM5α was discovered by virtue of the ability of the rhesus macaque ortholog to block replication of the human immunodeficiency virus 1 (HIV-1) in human cells (*33*). However, these more targeted interactions come at a cost, such that the antiviral range of such factors is often extremely limited. For instance, Fv1-B can

act on B-tropic Murine Leukemia Virus (B-MLV) but not against the highly related N-MLV, a difference in susceptibility that is governed by a single amino acid change between capsids (*26*). The highly tuned nature of these immune responses means that virus evasion can occur through the acquisition of mutations that reduce binding affinity of host proteins, rather than employing dedicated protein antagonists (**Figure 2B**).

The primary requirement for ISG inhibition of viral replication upon target recognition represents a general "rule of engagement" in host-virus interactions (*34*). In this way, ISGs can be considered "effector PRRs." These molecular recognition events allow for the functional classification of ISGs based on the dichotomy between the type of ISG substrate (ubiquitous/essential versus virus-specific) and the range of antiviral activity (broad or narrow). The sole outliers to this apparent dichotomy are the Mx1 GTPases. Both human MxA and mouse Mx proteins act broadly against a wide-spectrum of RNA and DNA viruses (*35*), suggesting that they must exploit some general molecular cue of infection. In contrast to this expectation, MxA instead appears to act via highly specific recognition of different viral proteins from diverse viruses. Hence, the Mx GTPases provide a unique opportunity to investigate the molecular basis underlying flexibility in target recognition and antiviral specificity.

*The discovery of Mx antiviral genes*

The first description of IFN by Isaacs and Lindenmann was a major landmark toward the acknowledgement and understanding of a genetic basis for host susceptibility to viral infection (*36*). In 1962, Lindenmann published a short report describing the unique resistance of the inbred mouse strain A2G to mouse-adapted myxovirus (influenza virus) strains (*37*). Crossing of A2G mice with susceptible mice yielded $F_1$, $F_2$ and backcrossed offspring with percentages of

susceptibility that were consistent with the presence of a single, dominant allele that conferred protection against otherwise lethal doses of mouse-adapted influenza virus in a heritable fashion (*38*). Owing to the phenotype, the symbol *Mx* (myxovirus-resistant) was designated for the putative allele. Initial attempts to overcome the resistance phenotype of *Mx*-bearing mice by various manipulations (e.g., cortisone therapy, neonatal thymectomy) failed (O. Haller, personal communication). Then in a wonderful example of serendipity - 22 years following the discovery of IFN - anti-IFN neutralizing antibodies were found to completely abrogate protection in *Mx*-bearing mice (*39*). This result was surprising at that time as prevailing theory suggested a non-specific affect for IFN in host immunity. This observation led to mapping of the myxovirus resistance 1 gene (*Mx1*), which beautifully demonstrated that *Mx* cDNA transduced into susceptible cells resulted in expression of a protein that provided resistance to influenza (*40*). These studies yielded the the first known instance of an IFN-regulated, single gene resistance to viral infection.

Around the same time, an interferon-induced human protein that cross-reacted with the anti-Mx1 C212 monoclonal antibody was identified (*41*). Hybridization of *Mx1* cDNA probes allowed for the isolation of two distinct cDNA clones, which encode the human proteins MxA and MxB (*42, 43*). Human MxA was found to be functionally analogous to mouse *Mx1*$^{-/-}$ cells, conferring a similar resistance to viral infection (*44*). In contrast, an antiviral function for MxB would not be revealed for another 24 years ((*45-47*), see Appendix 1). Constitutive expression of human MxA was also sufficient to confer resistance to influenza virus infection in *Mx1* null and *Ifnar*$^{-/-}$ mice (*48*), which to date remains the most substantial evidence that human MxA acts similarly to the rodent Mx1 as a single gene determinant of host susceptibility to virus infection.

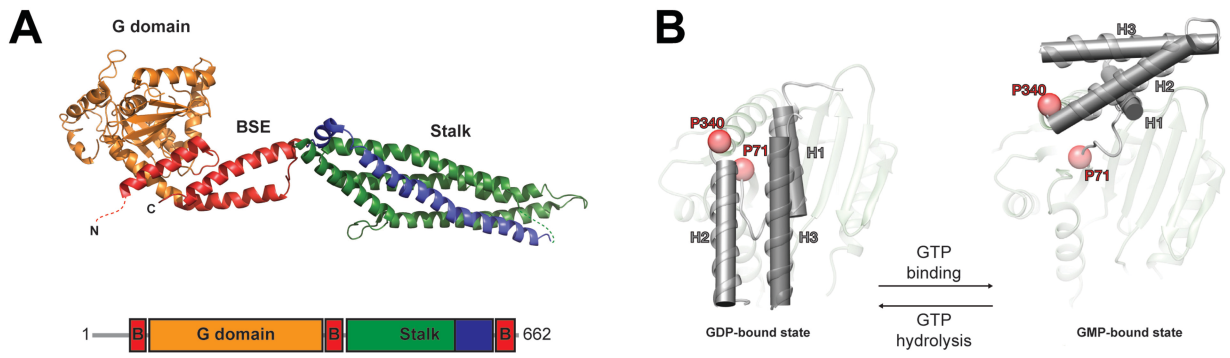



**Figure 3. Structural and evolutionary insights into MxA antiviral activity**

**A.** The crystal structure of human MxA (pdb 3SZR) (*49*) is depicted with the GTPase domain (G, orange), bundle-signaling element (BSE, red) and stalk (blue) oriented top to bottom. Disordered loops are represented as dashed lines, and have been manually drawn in using the software PyMol (*50*). **B.** Upon GTP hydrolysis the BSE of MxA undergoes a conformational change relative to the G domain, which leads to a "power stroke" movement of the stalk. **Figure 3B** was modified from (*51*) with permission, Elsevier license number 3591700700091.

*Mechanism of MxA action*

The human *Mx1* gene encodes a 662 amino acid (76 kDA) protein comprised of an amino (N)-terminal GTPase domain, middle domain and carboxy (C)-terminal GTPase effector domain (GED) (*52*). Phylogenetically, MxA proteins are most closely related to Dynamin and Dynamin-like GTPases (*53*). As such, MxA exhibits canonical Dynamin-like characteristics of low affinity for guanine nucleotides and high intrinsic rates of GTP hydrolysis, which is dose-responsive and dependent on oligomerization (*54*). However, an understanding of how the GTPase function contributes to MxA antiviral activity has been elusive.

Two crystal structures of the human MxA protein (*49, 54*) reveal an elongated three-domain protein comprised of an N-terminal globular GTPase-containing head (G domain) and a C-terminal helical stalk, which are connected by a hinge-like bundle-signaling element (BSE) (**Figure 3A**). The asymmetric unit revealed a domain-swapped orientation of two MxA monomers, permitting the identification of intermolecular contacts in the stalk and BSE domains. Point mutations that abolish homo-oligomerization also inhibit MxA antiviral activity. Although

MxA exists as a stable tetramer in solution, cryo-electron microscopy of both Dynamin (*55*) and MxA (*56*) indicate the presence of higher-order assemblies on lipid tubules in the presence of GTP. Computational modeling of MxA dimers predicts a stalk-mediated assembly into a ring-shaped antiviral complex, with G domains facing outward (*49*). Importantly, this model provides an explanation for oligomerization-dependent GTP hydrolysis, where self-propagating GTPase activity is dependent on higher-order assembly that brings G domains of neighboring tetramers in close proximity (*57-59*). Moreover, this orientation suggests that upon formation of the MxA antiviral complex GTP hydrolysis can coordinately signal through the BSE to the stalk. These inferences were corroborated by structures of GTP and GDP-bound "stalkless" MxA dimers, which indicate major structural changes in the BSE relative to the G domain. Together, these studies suggest that an MxA "power stroke" is the mechanism through which stored chemical energy is used to perform mechanical work on targeted structures (**Figure 3B**) (*51, 60*), highly analogous to Dynamin-mediated fission of associated membranes (*55*).

Together, the crystal structure of MxA and subsequent biochemical dissection have culminated in a compelling model for how the Dynamin-like GTPase protein architecture elicits antiviral activity (*51*). However, this model also highlights historical difficulties in understanding MxA target recognition. The structures reveal, for instance, that a complex network of protein-protein contacts govern GTP hydrolysis and oligomerization. Therefore, mutagenesis or truncation-based methods have typically resulted in non-specific loss-of-function. For example, established Mx mutants that lack antiviral activity (e.g., H630K in rat Mx2) can now be explained by the disruption of intramolecular contacts between the BSE and stalk (*49, 61*). Thus, although the crystallographic information has profoundly advanced our biochemical and

structural understanding of the MxA antiviral complex, they still leave unanswered the important question of how this complex might engage with its myriad viral targets.

*The MxA – NP interface is a critical molecular determinant of influenza virus zoonoses*

Given its historical origins, the molecular interactions between MxA and viral targets have been most fully described for *orthomyxoviruses*, and in particular for influenza A virus. Multiple lines of evidence suggest that the nucleoprotein (NP) is the primary viral determinant underlying susceptibility to MxA. For example, variation in the NP segment can completely explain the differential susceptibility to human MxA by the avian H5N1 (susceptible) and human H1N1 (resistant) influenza viruses (*62, 63*). These susceptibility determinants were recently mapped by comparing NPs from MxA-susceptible (H5N1) and -resistant (H1N1/1918 and H1N1/2009) influenza strains (*64*). Escape residues, which form a surface-exposed cluster in the body domain of NP, are sufficient to confer MxA-resistance when introduced into the otherwise susceptible H5N1 NP. Similar MxA-escape mutations were also shown to have accumulated in H3N2 viruses since its introduction into humans from swine. Interestingly, the emergent human H7N9 influenza virus and its avian precursor encode an NP containing a novel motif that renders it similarly resistant to human MxA, indicating that pre-adaptation to human MxA may have facilitated zoonotic transmission (*65*). Taken together, these studies implicate MxA as a major barrier to zoonotic transmission of influenza viruses, and that the MxA-NP interface is a molecular barrier to cross species transmission.

*The conundrum of MxA antiviral breadth despite target specificity*

The influence of MxA stretches far beyond *orthomyxoviruses*. Indeed, the breadth of MxA antiviral activity encompasses a striking diversity of both negative sense and positive sense RNA viruses, including *bunyaviruses*, *paramyxoviruses*, *picornaviruses*, *rhabdoviruses* and *togaviruses* (reviewed in (*35*)). Whether MxA similarly serves as a molecular barrier to cross-species transmission of these other viruses is unknown. However, constitutive expression of human MxA in *Ifnar*$^{-/-}$ *Mx* null mice is broadly protective (*48, 66*). The nature of MxA target recognition with other viruses has not been resolved to the same detail as that of *orthomyxoviruses*. Nonetheless an amazing diversity of viral targets has emerged. For example, infection by *Bunyaviruses* including La Crosse virus, Rift Valley fever virus and Bunyamwera virus results in co-localization of MxA and the viral nucleocapsid (N) protein at an ill-defined perinuclear compartment (*67*). MxA also interacts with the NP protein from the tick-borne *orthomyxovirus* Thogotovirus (THOV) (*68*). This suggests that similarly to the influenza virus NP, MxA engages the N protein of *bunyavirus*es and divergent *orthomyxoviruses* to elicit its antiviral function.

Consistent with MxA recognition of divergent viral substrates, MxA impedes viral replication in a virus-specific fashion. For example, MxA blocks nuclear import of THOV RNPs via direct interaction with NP (*68, 69*), whereas FLUAV replication is disrupted following RNP nuclear import and primary transcription (*70*). MxA also inhibits mRNA synthesis of the cytoplasmic replicating VSV (*71*). The incongruence of MxA action across virus families has led to a model in which MxA binding of viral targets sequesters some rate-limiting viral component or disrupts viral protein assemblies, despite ultimately affecting different stages of viral replication. Regardless of the molecular mechanism by which MxA effects restriction, MxA binding to viral targets is a clear requirement for antiviral function. One possibility is that MxA

may target proteins with similar functionality across divergent viruses. However, the existence of a pan-viral epitope for MxA targeting is difficult to reconcile with recent studies that extend MxA antiviral activity to DNA viruses. For example, MxA inhibits the *hepadnavirus* hepatitis B virus (HBV) (*72*) via the hepatitis B core antigen protein (HBcAg) (*73*). MxA has also been reported to restrict large double-stranded DNA viruses including the *orthopoxvirus* monkeypox (*74*) and the poxvirus-like *Asfarvirus* African swine fever virus (ASFV) (*75*); for these viruses MxA targets are unknown. MxA also restricts Semliki Forest virus (SeFV), a positive-sense RNA virus (*48, 76*). This restriction appears to be independent of SeFV structural proteins (including the genome organizing nucleocapsid), suggesting that MxA targets a component of the SFV replicase (*76*).

Taken together, the emerging model for MxA antiviral activity paradoxically requires that target recognition occur across highly divergent proteins from a diversity of viral families. This presents a molecular conundrum for MxA target recognition: how does MxA maintain the ability to recognize so many different viral proteins, each with the capacity to rapidly evolve to evade recognition?

*Pathogen-driven evolution reveals molecular surfaces critical to host-virus interactions*

Evolutionary signatures have been commonly used to gain insight into protein function. For example, active site residues essential to the chemistry performed by enzymes are strictly conserved over large phylogenetic distances. It is assumed that, given sufficient evolutionary time, random mutations have been tested by natural selection at nearly all possible positions (i.e., observed mutations are the result of selection or drift). Therefore, the conservation of genes or

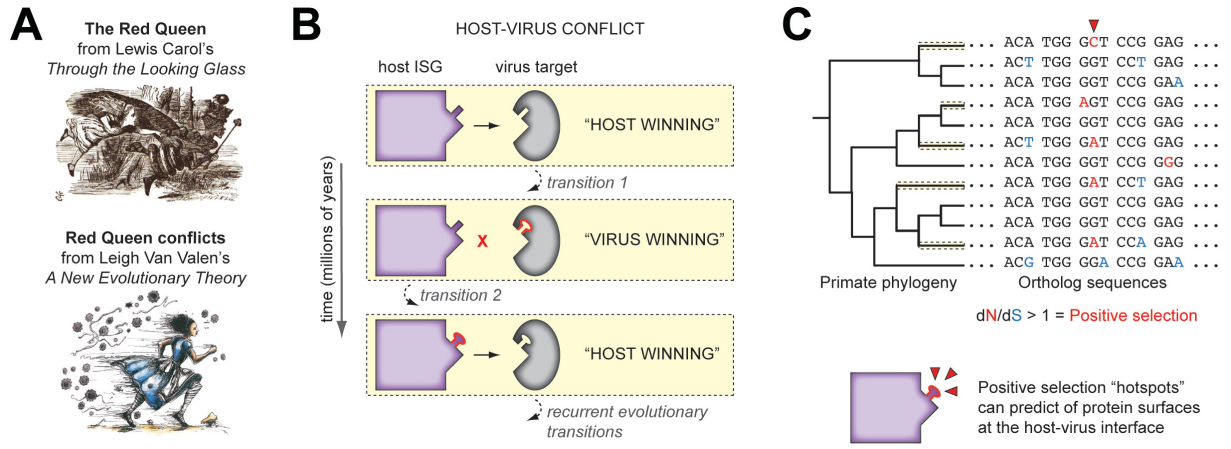



**Figure 4. Identification of critical antiviral interfaces using signatures of positive selection from Red Queen host-virus conflicts**

**A.** In Lewis Carol's *Through the Looking Glass* the Red Queen exclaims to Alice, "it takes all the running you can do, to keep in the same place." This concept was used by Leigh VanValen to describe his evolutionary theory on antagonistically evolving systems in which fitness is constant over time as a result of an ever-changing environment. **B.** Depicted is a generalized arms race between a host antiviral factor (purple) and its viral target (grey). Host recognition, which restricts viral replication, selects for variants in the viral population that evade host factor binding. This in turn selects for host variants that re-establish target interaction. The effective population size and mutation rate of viruses decreases the number of generations required to evolve adaptive mutations. Note that the 'direction' of the arms race can be reversed (e.g., viral antagonists that target broadly-acting antiviral factors). **C.** Positive selection (dN/dS > 1) can be assessed at the codon (shown) or whole-gene level in an alignment of orthologous genes. Independent, lineage-specific changes (yellow bars) to the same codon reflect the recurrent and episodic nature of host-virus conflicts. "Hotspots" of positive selection tend to cluster on surfaces of host antiviral proteins that directly engage viral components. Figure 4A (upper and lower) was reproduced from (*77*) and (*78*) respectively with permission, Nature Publishing Group license number 3591710543678 and Elsevier license number 3591711478918.

specific codons is presumed to come about because deleterious mutations have been negatively selected, or purified from the population.

Signatures of rapid evolution can be similarly used to gain substantial molecular insights into host-virus interactions. The antagonistic relationship between pathogenic viruses and their infected hosts drives the rapid evolution of both parties. Under this scenario, host and viral genomes present an ever-changing fitness landscape to the other, such that beneficial mutations provide only a temporary advantage. Such cycles of host-virus adaptation and counter-adaptation exemplify classical "Red Queen" genetic conflicts (**Figure 4A**) (*79*). Because the relative fitness of either party is in direct opposition to the other, one side is always losing. Therefore, host-virus

conflicts are under constant evolutionary pressure to innovate by positive (i.e., diversifying) selection (**Figure 4B**).

A common metric used to evaluate selection is the dN/dS statistic, which calculates the observed rate of non-synonymous (dN) relative to synonymous (dS) nucleotide substitutions in an alignment of orthologous gene sequences. For the example of enzyme active site residues, dN/dS would be predicted to be significantly less than the neutral expectation (dN/dS = 1.0) owing to an underrepresentation of culled non-synonymous changes. In contrast, immunity genes are among the fastest evolving class of genes in primate genomes (*80*), due to constant selection for innovation in host-virus conflicts. dN/dS-based inference of natural selection on protein-coding genes ignores phenotypes that manifest as the result of non-equivalent synonymous mutations. For example, RNA secondary structure (*81*), codon bias (*82*) and "duons" (*83*) constrain the evolution of protein-coding genes, and may artificially inflate dN/dS ratios focused solely at the codon level. This weakness notwithstanding, dN/dS is a powerful tool that has yielded substantial evolutionary and molecular insights into host-virus interactions (for example, (*84-101*)).

Signatures of positive selection in antiviral proteins are often concentrated on surfaces that target (or are targeted by) viruses. Because independent, adaptive mutations tend to converge on similar means to achieve the greatest phenotypic advantage, a recurrent pattern emerges at host-virus interfaces. As a result, clusters of positively selected sites, or "hotspots," can be used to predict residues that significantly impact the affinity of host-viral interactions (**Figure 4C**) (*34*). Therefore, like purifying selection denotes conserved protein function, evolutionary signatures of positive selection can guide the identification of critical host-virus interfaces.

My graduate thesis uses evolutionary signatures of host-virus conflicts to gain a molecular understanding of MxA antiviral breadth. This evolution-guided approach has identified rapidly evolving target recognition elements that mediate MxA antiviral specificity, and reveals general principles of host-virus interactions that are critical to host susceptibility to viral infection and the molecular determinants of viral host range.

## Chapter 2: Methods

Parts of this Chapter were modified from (*1*) and (*102*) with permission, Elsevier license number 3591720358649 and 3591720484968.

### *Cell culture*

Mammalian cell lines, including 293T, BSC40, CRFK and Vero were maintained in DMEM (Gibco) supplemented with 10% FBS and 5% penicillin/streptomycin. BSC40 and CRFK Mx-expressing stable cell lines were maintained under 10.0 or 3.0 ug/mL puromycin selection.

### *Amplification of primate MxA genes*

RNA was isolated using the Qiagen RNeasy kit from cell lines obtained from Coriell Cell Repositories (Camden, NJ), as described in **Table 1**. Primate MxA genes were amplified using one-step reverse-transcription polymerase chain reaction (RT-PCR) was conducted using SuperScript III One-Step RT-PCR with Platinum Taq (Invitrogen) to produce complementary DNA (cDNA) using degenerate primers:

F: 5’ – CAAAGAAGGAAGATGGTTSTTTCCGAAGTGG - 3’

R: 5’ – TTAACCGGGGAACTGGGCRAG – 3’

cDNA was Sanger sequenced and assembled using CodonCode Aligner. The following sequences were obtained from Ensembl and were not independently validated: *M. mulatta* (rhesus macaque): ENSMMUT00000021494 and *C. jacchus* (marmoset): ENSCJAT00000021974.

*MxA sequence analysis*

Sequences were aligned in ClustalX and edited to remove indels. ML tests were performed with CODEML using the PAML software suite (*103*), as previously described (*84*). Briefly, sequences were subjected to ML tests using NS sites models disallowing (M7) or allowing (M8) positive selection. For each comparison the models allowing positive selection gave the best fit to the data. The result was consistent under varying models of codon frequency (F61 and F3x4). The same site-specific analyses also provided individual amino acids with high posterior probabilities that are consistent with positive selection.

*Plasmids and stable cell line production*

MxA coding sequence was amplified using AccuPrime Pfx SuperMix (Invitrogen) from cDNA derived from indicated primate species and subcloned into either pcDNA3 with an N-terminal hemagglutinin (HA) or Flag epitope for transient expression studies, or pQCXIP with a C-terminal 3xFlag epitope for the generation of stable cell line. Mx chimeras were generated by PCR. Point mutations were generated using Quikchange site-directed mutagenesis (Stratagene). The HIV-1-based gagpol expression construct JK3, pCMV-tat (fortransactivation of the HIV long terminal repeat of JK3) and pCMV-VSV-g were used to generate single-round infectious virus for the production of Mx-expressing CRFK and BSC40 stable cell lines.

*THOV minireplicon assay*

The THOV minireplicon assay was performed in 293T cells in a 96-well format. Briefly, 4.0 ng each of PB2, PB1 and PA, 1.0 ng of NP, all in pCAGGS expression vector, as well as 20 ng pHH21-vNP-FF-Luc (firefly luciferase), 50 ng Tk-luc (Renilla) (Promega) and varying input

or a fixed amount of 200 ng MxA plasmids were co-transfected into 293T cells using the reagent TransIT-LT1 (Mirus Bio). At 24h post-transfection, luciferase activity was measured using the Dual-Glo system (Promega). Expression of MxA proteins was detected by Western blot analysis using mouse anti-HA.11 monoclonal antibody 16B12 (Sigma) and rabbit anti-βactin polyclonal antibody AB-8227 (Abcam). Corresponding data from minireplicon assays and Western blots are derived from samples generated from a single master mix. All experiments were done in triplicate.

*Co-immunoprecipitation*

For immunoprecipitation, 293T cells were transfected in six-well plates with 1 μg of Flag-MxA-encoding plasmids for 24 h and then infected with 10 moi of THOV for additional 24 h. Cells were lysed in buffer, 50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40. The supernatants were used for immunoprecipitation of Flag-MxA using Anti-FLAG-M2 affinity gel (Sigma) for 2 h at 4°C. After washing in lysis buffer the precipitated proteins were eluted in SDS-sample buffer at 95°C for 5 min. Co-precipitated MxA and viral NP were detected by Western blotting using monoclonal antibodies specific for MxA and THOV NP (*68*) and β-tubulin (Sigma).

*Viruses and viral infection*

Vero cells were transfected in 24-well plates with 250 ng MxA-expressing plasmids and nanofectin (PAA) for 24 h and then infected with 10 moi of THOV (SiAr126 strain) for 24 h or 5 moi of influenza A virus (A/Thailand/1/04 strain) for 5 h. Infected cells were fixed with 3%

paraformaldehyde and MxA and viral NP expression were detected by immunofluorescence analysis using specific antibodies as described previously (*63, 104*).

VSIV (Indiana strain) was a kind gift from John Rose. MxA-expressing BSC40 cells were plated in 24-well format and infected at a 0.01 m.o.i. (unless otherwise indicated) for 1 h. Infected cells were washed in PBS and supplied fresh growth media. 24h-post infection virus-containing supernatants were harvested and titered on BSC40 cells by plaque assay. For experimental evolution studies, human MxA-expressing CRFK cells were initially infected at 0.01 m.o.i. followed by blind passaging with titering every fifth passage. Passage 15 viruses were plaque purified from agar overlays, expanded on human MxA-expressing CRFK cells and titered on BSC40 cells by plaque assay.

Vaccinia virus (Western reserve strain) was a kind gift from Rich Condit. VACV infections of MxA-expressing BSC40 cells were done in 24-well format at a 0.3 m.o.i. 48h-post infection, cells were frozen and virus-containing supernatants were clarified by centrifugation at 500 x *g*. Supernatants were titered on BSC40 cells by plaque assay.

*Proximity ligation assay*

MxA-expressing BSC40 cells were grown on glass coverslips in 6-well format, and infected with VSIV at an m.o.i. 5.0 for 5h. Cells were washed in PBS and fixed in 4% paraformaldehyde. Fixed cells were permeabilised in methanol at -20$^{o}$C, blocked in 4% FBS, 0.1% Tween-20, PBS for 1h, and stained using rabbit anti-FLAG (Sigma) and mouse anti-VSIV-N 10G4 or anti-VSIV-M 23H12 (KeraFAST) for 1h in a humidified chamber at 37$^{o}$C. All Duolink PLA reagents were purchased from Sigma. PLA probe incubation, ligation and rolling circle amplification were carried out following the manufacturer's protocol directly on

coverslips. Coverslips were mounted using ProLong Gold Antifade Mount with DAPI (Molecular Probes) and visualized on a Leica TCS SP5 II confocal microscope with image acquisition using LASAF software.

*Sequence deposition*

All novel primate MxA sequences are deposited into Genbank under accession numbers JX297228-JX297248.

**Table 1. Source of primate MxA cDNAs**

| Species | Common Name | Sample source | GenBank Accession |
|---|---|---|---|
| *Macaca fasicularis* | crab-eating macaque | M05221[#] | |
| *Macaca mulatta* | rhesus macaque | | NP_001073161 |
| *Macaca sylvanus* | Barbary macaque | GB72(*105*) | |
| *Cercocebus atys* | sooty mangabey | G077[##] | |
| *Chlorocebus pygerythrus* | African green monkey | CRL-1651 (ATCC) | |
| *Miopithecus talapoin* | talapoin monkey | PR00716 (Coriell) | |
| *Trachypithecus francoisi* | Francois' leaf monkey | PR01099 (Coriell) | |
| *Colobus guereza* | colobus monkey | PR00980 (Coriell) | |
| *Homo sapiens* | human | GM10969 (Coriell) | |
| *Pan paniscus* | bonobo | AG05253 (Coriell) | |
| *Pan troglodytes* | chimpanzee | AG06939 (Coriell) | |
| *Gorilla gorilla* | gorilla | AG05251 (Coriell) | |
| *Pongo pygmaeus* | orangutan | AG05252 (Coriell) | |
| *Hylobates agilis* | agile siamang | PR00773 (Coriell) | |
| *Symphalangus syndactylus* | island siamang | PR00722 (Coriell) | |
| *Nomascus leucogenys* | white-cheeked gibbon | PR01037 (Coriell) | |
| *Callithrix jacchus* | common marmoset | | XM_002761438 |
| *Saguinus midas* | red-handed tamarin | PR00550 (Coriell) | |
| *Aotus trivirgatus* | three-striped night monkey | CRL-1556 (ATCC) | |
| *Saimiri sciureus* | common squirrel monkey | AG05311 (Coriell) | |
| *Alouatta sara* | Bolivian red howler monkey | PR00708 (Coriell) | |
| *Lagothrix lagotricha* | common woolly monkey | AG05356 (Coriell) | |
| *Callicebus moloch* | dusky titi | AG06115 (Coriell) | |
| *Pithecia pithecia* | white-faced saki | PR00239 (Coriell) | |

#Identifier from the Washington National Primate Research Center
[##]Identifier from the Tulane National Primate Research Center

## Chapter 3. Evolution-guided identification of antiviral specificity determinants in the broadly acting innate immunity factor MxA

Parts of this Chapter were modified from (*102*) with permission, Elsevier license number 3591720484968.

### Abstract

**Human myxovirus resistance protein, MxA, is an important effector of the interferon-induced innate immune response to a wide diversity of viral families. MxA executes its antiviral action through specific recognition of divergent viral structures. However, the basis of MxA antiviral specificity is not well understood. We used an evolution-guided approach to identify the loop L4 of MxA as a hotspot for recurrent positive selection in primates. We show that single amino acid changes in L4 are necessary and sufficient to explain dramatic differences in species-specific antiviral activity of primate MxA proteins against the *orthomyxoviruses* Thogoto virus and influenza A virus. Taken together, our findings identify a genetic determinant of MxA target recognition and suggest a model by which MxA achieves antiviral breadth without compromising viral specificity.**

**Introduction**

MxA is a dynamin-like GTPase with antiviral activity against a wide range of RNA and DNA viruses (*35*). The antiviral breadth exhibited by MxA is remarkable because it hinges upon detection of unique viral structures that differ across virus families. For example, differences in resistance and susceptibility between avian (MxA-sensitive) and human (MxA-resistant) influenza virus isolates have been shown to be solely dependent on differences in the nucleoprotein (NP) (*62*). Furthermore, MxA activity against the *alphavirus* Semliki Forest virus (SFV) is independent of SFV NP or other structural proteins (*76*), and MxA antiviral activity against DNA viruses like hepatitis B virus (HBV) (*73*) and African swine fever virus (ASFV) (*75*) is dependent on unique viral components. However, the evolutionary and molecular basis for MxA antiviral specificity is unknown.

The recently solved crystal structure of human MxA (hsMxA) highlights the difficulty in pinpointing molecular determinants of MxA antiviral specificity. MxA structure resembles that of other members of the dynamin-like large GTPase superfamily (*49*), consisting of a N-terminal GTPase domain (G domain) and a C-terminal stalk. These two structural domains are linked by a bundle-signaling element (BSE) that is necessary to transfer structural changes during GTP binding and hydrolysis to the stalk structure. Nonetheless, while the crystal structure of MxA illustrated the coupling between different domains, it did not reveal a molecular basis for the long-standing problem of MxA antiviral specificity

In order to determine the basis of MxA specificity for viral recognition, we took an evolution-guided functional approach that capitalizes on the antagonistic arms race between MxA and its viral targets, and the genomic signatures it leaves on host genomes. The rapid accumulation of amino acid replacement changes (dN) relative to synonymous changes (dS), or

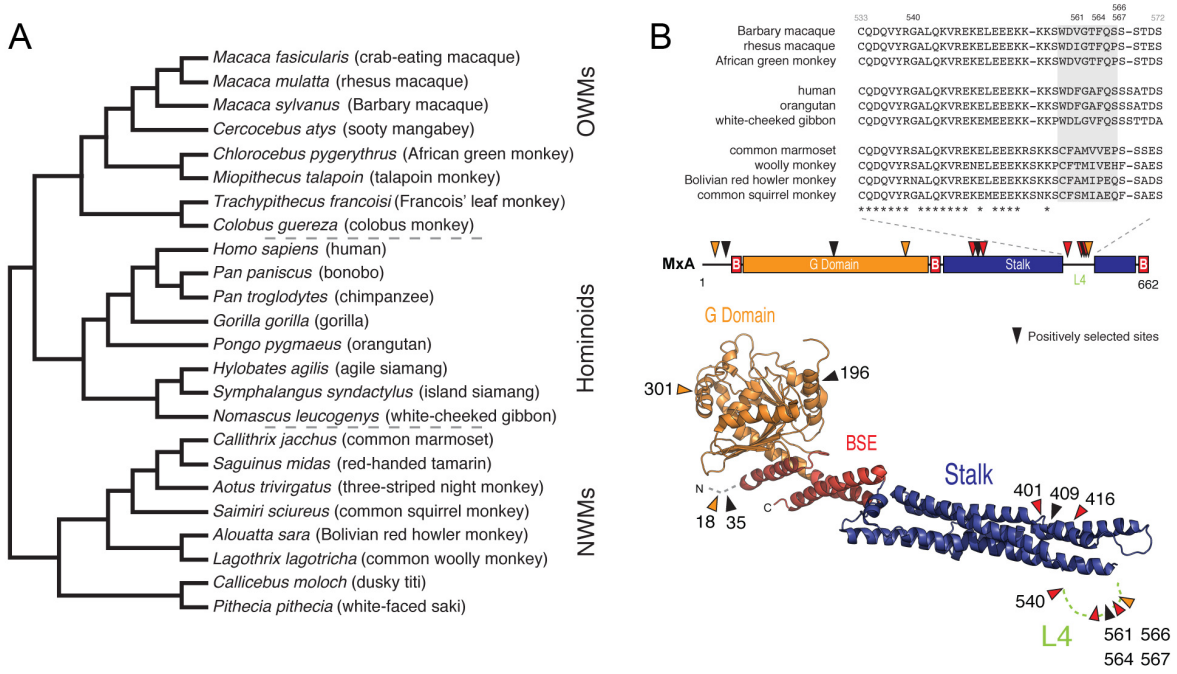
positive selection (dN/dS > 1), is one hallmark of antagonistic evolution between host and viral genomes. Because positive selection underlies phenotypic adaptation at antiviral interfaces (*34, 84*), we predict that the hypothetical viral substrate interface(s) of MxA would have evolved with the strongest signature of positive selection.

Here we show that the surface-exposed loop L4, which protrudes from the compact structure of the MxA stalk, bears such a signature of recurrent positive selection. We demonstrate that genetic variation in L4 of primate MxA is a major determinant of its species-specific antiviral activity against the *orthomyxoviruses* Thogoto virus (THOV) and an avian influenza A virus, and a single positively selected amino acid in L4 is sufficient to alter the antiviral specificity of primate MxA. Moreover, we show that hsMxA L4 can function as a modular determinant of antiviral specificity in the context of the highly divergent mouse ortholog to confer antiviral protection. We propose that, while the architecture of Mx proteins has been evolutionarily conserved, the antiviral specificity determinants have been subject to recurrent arms races with viral substrates. Taken together, our findings identify L4 as a genetic determinant of MxA antiviral specificity, and highlight the power of evolutionary methodologies to delineate host-virus interactions.

## Results

### *Rapid evolution of the unstructured loop L4 in primate MxA*

To evaluate how the antiviral specificity of MxA has been shaped during the evolution of primates, we amplified, sequenced and aligned *MxA* coding regions from 24 primate species representing ~43 million years of evolutionary divergence (**Figure 5A, Figure 11 and Table 1**).




| Whole gene analysis | P-value (M7 v. M8) | p1 (% sites)[i] | ω (dN/dS)[ii] |
|---|---|---|---|
| MxA (aa 1-662)[iv] | < 0.0001 | 3.7 | 4.31 |
| MxA (aa 1-662), excluding NWM | < 0.0001 | 4.4 | 3.82 |

| Domain analysis | P-value (M7 v. M8) | p1 (% sites) | ω (dN/dS)[ii] | Adaptively evolving amino acids[iii] |
|---|---|---|---|---|
| G-domain (aa 69-340) | 0.0177 | 5.1 | 2.32 | Y196 (0.924), I301 (0.985*) |
| BSE (aa 44-64, 341-360, 633-662) | 0.999 | 0.7 | 1.66 | --- |
| Stalk (aa 367-632) | < 0.0001 | 4.5 | 5.48 | I401(0.996**), H409 (0.925), T416 (0.999**) |
| L4 (aa 533-572) | < 0.0001 | 18.6 | 5.08 | G540 (0.997**), F561 (1.000**), F564 (0.941), S566 (0.999**), S567 (0.957*) |
| StalkΔL4[v] | 0.0003 | 2.6 | 4.82 | --- |

**Figure 5. Evolution of primate MxA**

**A.** The phylogeny was built using orthologous MxA sequences from 24 primate species (Figure 11 and Table 1), which is congruent with a previously published primate phylogeny (*106*). **B.** The loop L4 (amino acids 533-572) protein sequence alignment is shown from representative primates (top). Specific codons found to be evolving under positive selection in primate MxA are indicated in a linear schematic (middle) or a model of the MxA crystal structure (bottom) (*49*), where sites with a posterior probability greater than 0.95 and 0.99 are highlighted in orange and red, respectively. In the linear structure "B" depicts the three parts of MxA that form the bundle-signaling element (BSE). **C.** Evolutionary analysis for positive selection in primate MxA. *P* values generated from maximum likelihood ratio tests that fit the data to models of neutral (M7) or adaptive (M8) evolution are given for the entire MxA coding sequence (including or excluding new world monkey (NWM) sequences) or specific domains of the protein as indicated. *i.* Percentage of codons evolving under positive selection (p1). *ii.* Associated average dN/dS (ω) of p1 sites. *iii.* Codons with a high posterior probability (PP > 0.90) that supports the likelihood of a site having a dN/dS > 1 are presented with the PP for each residue in parentheses. Amino acids refer to residues found at indicated positions in hsMxA. *iv.* Indels removed from alignment. *v.* StalkΔL4 refers to an analysis of the stalk with amino acids 533-572 removed.

A phylogenetic tree generated from these *MxA* sequences was congruent with a previously published primate phylogeny (*106*) confirming that the *MxA* sequences obtained were orthologous (**Figure 5A**). Using this dataset, we calculated an average dN/dS value of 0.32 for *MxA*, which is consistent with purifying selection having acted on the majority of the *MxA* gene over primate evolution. However, an average dN/dS signature can obscure recurrent positive selection that might have acted on particular codons or protein domains. Indeed, when we compared maximum likelihood codon-based models that disallow positive selection (model M7) to those that allow dN/dS to exceed 1 (model M8), we found strong evidence that *MxA* has recurrently evolved under positive selection in primates (**Figure 5C**). While only a few *MxA* codons (3.7%) have evolved under positive selection, the signature of positive selection associated with them is quite strong (average dN/dS ratio of 4.31). These estimates suggest that although most residues in MxA have evolved under purifying selection to maintain its structure and enzymatic properties, positive selection has acted on MxA at discrete residues likely in response to antagonistic evolution with viruses.

We found no evidence for positively selected sites in the BSE domain of MxA, and limited evidence in the G and N-terminal domains (**Figure 5B and 5C**). Instead, the majority of positively selected sites were located in the MxA stalk domain, which has the most significant domain-specific signature of positive selection. Within the stalk domain by far the most striking enrichment of positively selected sites is in the unstructured loop L4 with 18.6% of codons evolving with an average dN/dS value of 5.08, revealing a dramatic history of positive selection in L4 across primates. The overall signal of positive selection in the stalk domain is markedly reduced when L4 residues 533-572 are removed from the analysis (**Figure 5C**). Interestingly, we observed that there has been an almost complete replacement of the L4 residues between 559 and

566 (**Figure 1B**, grey box) that distinguish hominoid and old world monkey from new world monkey sequences. However, the new world monkey MxA divergence is not the only cause of a positive selection signature; evolutionary analysis of only hominoid and old world monkey MxA sequences also showed robust evidence of positive selection in *MxA* even in this less divergent sequence set (**Figure 1C**). These analyses indicate that significant evolutionary pressures have acted on MxA throughout primate evolution. The cumulative evidence for positive selection in MxA, particularly given the enrichment of sites in L4, strongly suggests that this surface-exposed loop, which is disordered in the recently determined MxA structure (*49*), has been subject to recurrent and strong positive selection during primate evolution.

*Loop L4 is the determinant of MxA antiviral specificity against Thogoto virus (THOV)*

To determine whether the observed positive selection in primate MxA has functional consequences, *MxA* coding sequences from 15 representative primate species were subcloned into mammalian expression constructs with a N-terminal HA-epitope. Protein expression was measured by Western blot analysis of cell lysates from transfected 293T cells, and only minor variation in protein expression was observed across the 15 orthologs and hsMxA(T103A), an inactive form carrying a threonine to alanine mutation in the G domain (*107*) (**Figure 2A**). We then tested the antiviral function of primate MxA orthologs using the THOV minireplicon system as a readout of MxA antiviral activity. Previous studies have shown that hsMxA severely

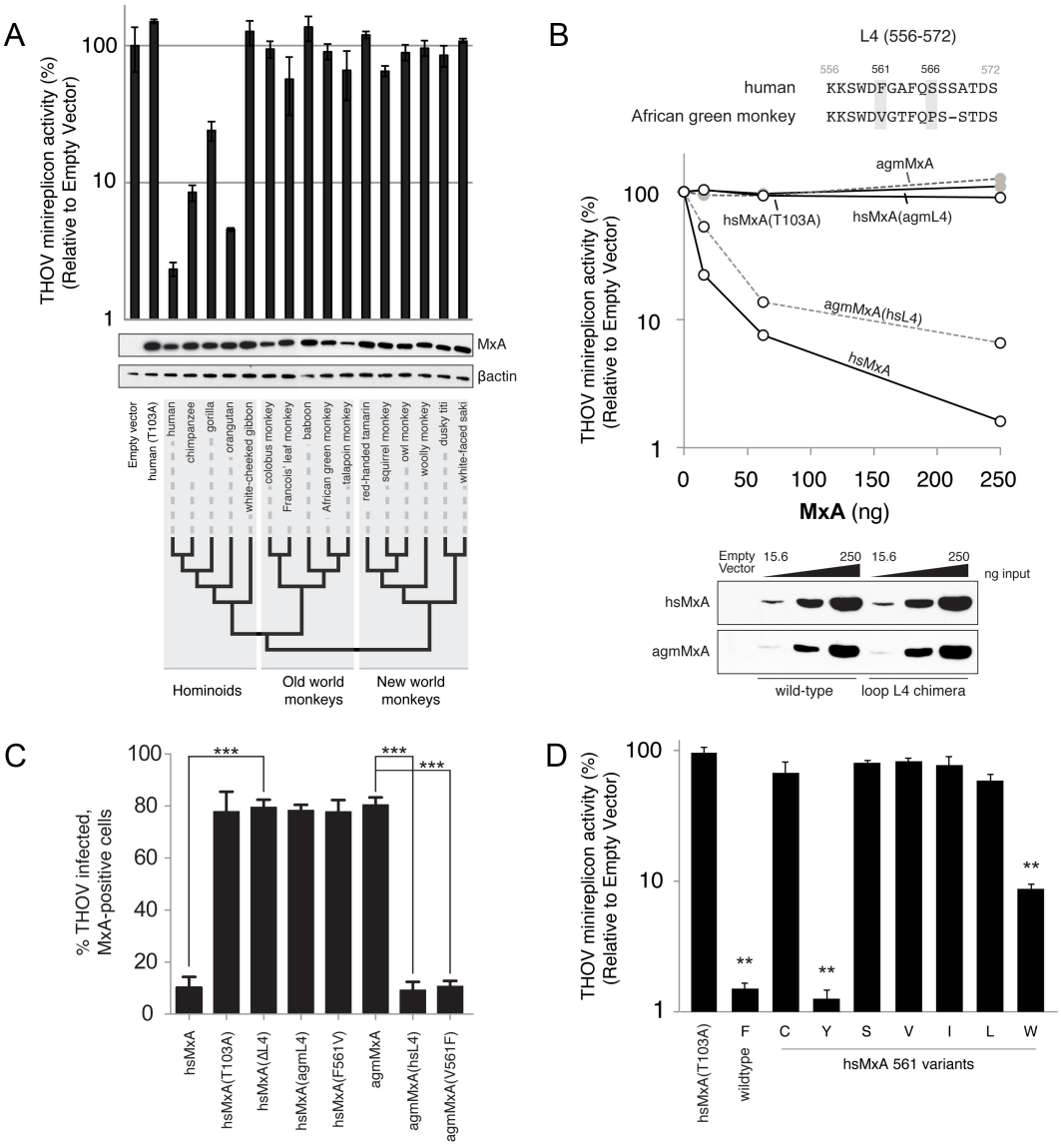



**Figure 6. Species-specific antiviral activity results from sequence divergence in loop L4 of primate MxA**

**A.** THOV minireplicon activity (luciferase levels) was measured 24h post-transfection, and was reported relative to levels observed in the absence of MxA. hsMxA(T103A) is an antivirally inactive mutant of hsMxA. Error bars represent standard deviation across three biological replicates. HA-tagged MxA and βactin protein levels were detected in lysates by Western blot. The phylogenetic relationships of MxA derived from representative species within hominoid, old world monkey and new world monkey lineages are indicated by a cladogram. **B.** Differential antiviral activity of human and AGM MxA orthologs is dependent on loop L4. The loop L4 (amino acids 556-572) protein sequence alignment is shown with positively selected sites that differ between hsMxA and agmMxA in grey (top). Dose-responsiveness of the antiviral activity of wild-type MxA and L4 chimeras was determined by co-transfecting increasing amounts of MxA expression constructs with the THOV minireplicon into 293T cells. MxA antiviral activity is measured as described in **Figure 6A** (middle). The Western blot analysis depicts the expression levels of the recombinant MxA proteins in the cell lysates (bottom). **C.** Restriction of THOV infectivity by human and AGM MxA is contingent on a single amino acid. Data are presented as percent THOV infected, Mx-positive cells as measured by immunofluorescence for THOV NP and MxA. Error bars represent standard deviation of three biological replicates. ***, $p<0.0001$ (t-test). **D.** Antiviral activity of hsMxA 561 variants. Data are presented as described in **Figure 6A**. **, $p<0.001$ (t-test) for values compared to hsMxA(T103A).

attenuates the THOV polymerase, and that its activity is sensitive to titration of both hsMxA and NP (*108*), making it an ideal 'surrogate' to test the functional consequences of variation across primate MxA orthologs. In this assay, the viral polymerase components (PB2, PB1, PA) and NP are delivered *in trans* to drive the expression of a firefly luciferase from a viral RNA minigenome in antisense orientation that is flanked by viral 5' and 3' UTRs. Firefly luciferase expression is detected 24h-post transfection and normalized to Renilla luciferase levels, serving as a transfection control. Luciferase expression in the absence of MxA is normalized to 100% polymerase activity.

We found that hsMxA reduces THOV minireplicon activity by ~50-fold when compared to either the inactive mutant hsMxA(T103A) or no MxA (**Figure 6A**). MxA from great apes reduced THOV primary transcription by 4 to 20-fold. In contrast, MxA proteins from gibbons, old world monkeys and new world monkeys were inactive (less than 2-fold reduction). The observed species-specificity indicates that genetic divergence of primate MxA orthologs bears dramatic functional consequences for MxA antiviral activity against THOV. Since there is a marked enrichment of positively selected sites in L4, we tested whether divergence in L4 might explain differences in antiviral activity. Therefore, we generated L4 chimeras between hsMxA (active) and the inactive ortholog from African green monkey (agmMxA). Introduction of human L4 into agmMxA (agmMxA(hsL4)) resulted in a dose-dependent rescue of activity against the THOV minireplicon (**Figure 6B**). Likewise, replacement of hsL4 with that of agmMxA (hsMxA(agmL4)) had the opposite effect, attenuating the block on THOV transcription. Thus, the loop L4, which we identified as evolving under positive selection, is a determinant for the antiviral specificity of MxA. Next, we tested the hsMxA and agmMxA chimeras in the context of virus infection. Vero (AGM) cells transfected with wildtype hsMxA or agmMxA were

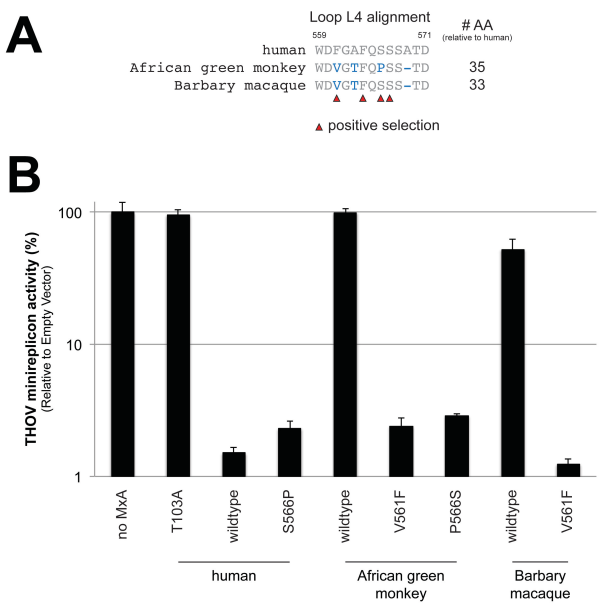



**Figure 7. Adjacent positively selected sites in the loop L4 do not affect antiviral specificity for THOV**

**A.** An alignment of the protein sequence of the loop L4 (amino acids 556-572) from hsMxA, agmMxA and sylMxA. **B.** THOV minireplicon activity (luciferase levels) was measured 24h post-transfection, and was reported relative to levels observed in the absence of MxA (*i.e.*, 100% activity in the empty vector control). hsMxA(T103A) is an antivirally inactive mutant of hsMxA. Error bars represent standard deviation across three biological replicates (bottom). sylMxA, MxA from Barbary macaque (*Macaca sylvanus*).

infected with THOV. We assessed intracellular viral replication by monitoring the accumulation of viral NP via immunofluorescence. In agreement with the results from the THOV minireplicon, viral infectivity was severely reduced in the presence of hsMxA and agmMxA(hsL4), but not agmMxA and hsMxA(agmL4) (**Figure 6C**). Additionally, the congruence of antiviral effects seen in the minireplicon assay in human cells (**Figure 6B**) and viral infection assay in AGM cells (**Figure 6C**) discounts the formal possibility that L4 effects are due to co-evolution with host-specific cofactors that differ between species. Taken together, our results demonstrate that the difference in antiviral function between hsMxA and agmMxA is the result of sequence variation in L4.

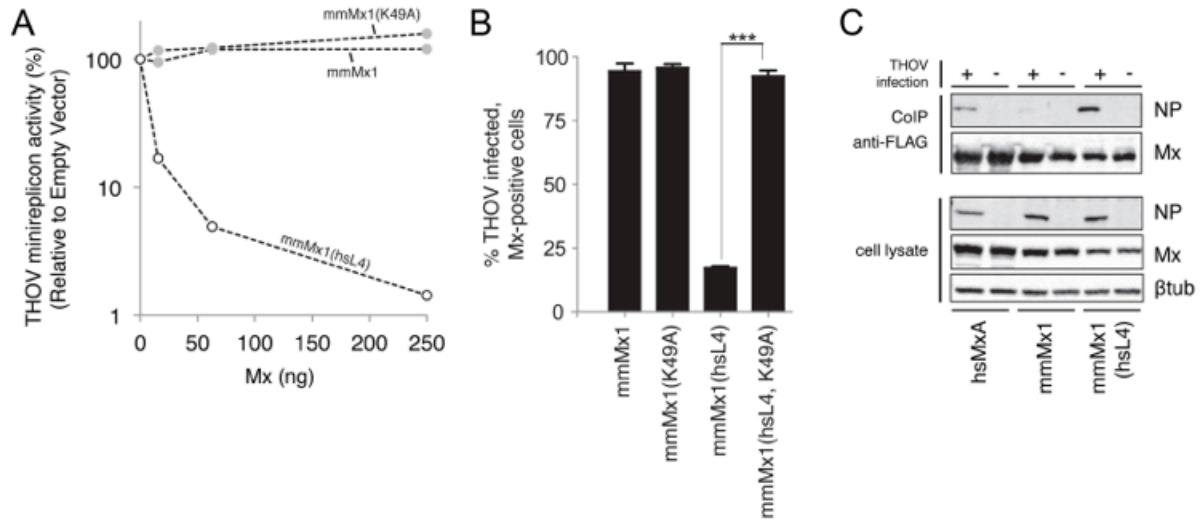



**Figure 8. Loop L4 is a functional module of MxA antiviral activity**

**A.** Human L4 rescues mouse Mx1 antiviral activity against the THOV minireplicon. mmMx1 expression constructs were co-transfected with the THOV minireplicon system into 293T cells; the experiment was otherwise carried out as in **Figure 6B**. mmMx1(hsL4) contains an exchange of mmL4 for the L4 of hsMxA. mmMx1(K49A) is a GTPase inactive mutant with no antiviral activity. **B.** Human L4 determines the antiviral activity of mmMx1 against THOV infection. Percent infectivity was measured as described in **Figure 6C**. Error bars represent standard deviation of three biological replicates. ***, $p<0.0001$. **C.** L4 mediates the interaction between MxA and THOV NP. Co-immunoprecipitation of viral NP with hsMxA and mmMx1 proteins was performed with lysates of THOV-infected (+) or uninfected (-) 293T cells transfected with the indicated Flag-tagged Mx expression constructs. Immunoprecipitation was performed using anti-Flag-affinity gel and the precipitated proteins were detected by Western blot with viral NP and MxA-specific antibodies. The lower panel shows the input of viral NP, MxA and β-tubulin in whole cell lysates.

*A single amino acid confers MxA antiviral specificity for THOV*

HsMxA and agmMxA differ by four amino acids in L4. Of these, we found that two positions, 561 and 566, are evolving under recurrent positive selection (**Figure 5B and C**). We focused on an analysis of position 561 since the serine residue at position 566 is shared between hsMxA and orthologous MxA proteins inactive against THOV (e.g. Barbary macaque, **Figure 7**). We mutated the phenylalanine at position 561 in hsMxA to the valine found in agmMxA, and also introduced the reciprocal V561F mutation into agmMxA. These single amino acid changes completely recapitulated the L4 phenotype in both hsMxA and agmMxA backgrounds in infected cells and in the THOV minireplicon system (**Figure 6C and Figure 7**). Thus, a single

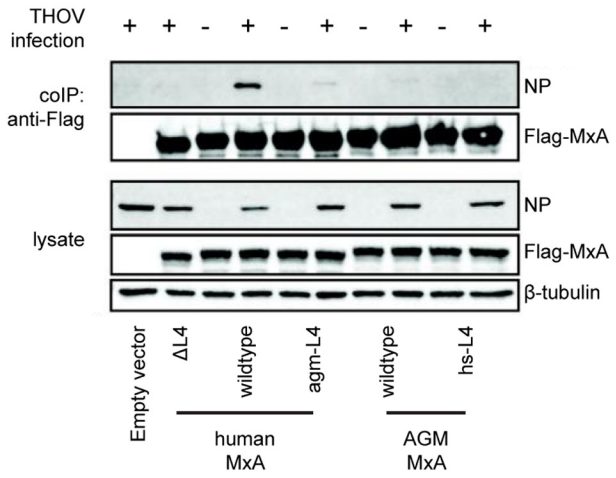



**Figure 9. Non-L4 determinants influence the interaction between primate MxA and THOV NP**

Co-immunoprecipitation of viral NP with hsMxA and agmMxA was performed with lysates of THOV-infected (+) or uninfected (-) 293T cells transfected with the indicated Flag-tagged MxA expression constructs. Immunoprecipitation was performed using anti-Flag-affinity gel and the precipitated proteins were detected by Western blotting using viral NP and MxA-specific antibodies. The lower panel shows the input of viral NP, MxA and b-tubulin in whole cell lysates.

amino acid evolving under positive selection in L4 largely determines viral specificity of MxA for THOV.

To understand in greater detail the molecular consequences of changes at residue 561, we tested amino acids present in hominoid and old world monkey species (hsMxA -F561I, -F561L and -F561V) as well as those that could be sampled by a single bp mutation of the hsMxA F561 codon (hsMxA -F561C, -F561Y and -F561S). Only wildtype F561 and F561Y restricted the THOV minireplicon whereas the other mutations dramatically reduced the antiviral activity of hsMxA (**Figure 6D**). Given the biochemical similarities of phenylalanine and tyrosine, we tested hsMxA-F561W, which also restricted THOV, albeit to a lesser degree (**Figure 6D**). These results reinforce the concept that MxA antiviral specificity for THOV is conferred by a single amino acid with specific biochemical requirements.

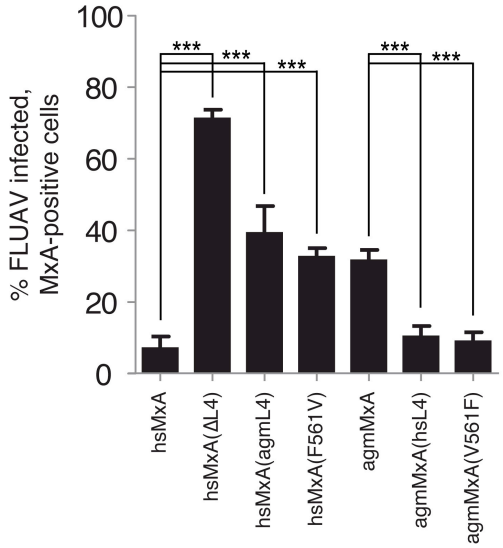



**Figure 10. The loop L4 mediates functional differences in primate MxA antiviral activity against influenza A virus**

Restriction of A/Thailand/1/04 infection by hsMxA and agmMxA. Data are presented as percent of infected, Mx-positive cells as measured by immunofluorescence for viral NP and MxA. Error bars represent standard deviation of three biological replicates. ***, $p<0.0001$ (t-test).

*Functional conservation of L4*

One prediction of the proposed role for L4 in MxA target recognition is that L4-mediated specificity is modular. If that is the case, the antiviral properties of hsMxA might be transferable to a more distantly related Mx protein simply by transplanting the loop L4 of hsMxA. Mouse Mx1 (mmMx1) is relatively inactive against THOV although it does confer some protection *in vivo* (*104*). We made chimeras replacing the loop L4 of mmMx1 with that from hsMxA, mmMx1(hsL4), and compared it to wildtype mmMx1 in the THOV minireplicon assay. Catalytically-deficient chimeras were also constructed by mutating the K49A in the G domain (*109*). We found that mmMx1(hsL4) gained robust activity against THOV (**Figure 8A**). We repeated these experiments using THOV infection of transfected Vero cells and again found that mmMx1(hsL4) is significantly more potent in blocking THOV than mmMx1 (**Figure 8B**). Thus,

the loop L4 encodes mmMx1 specificity for THOV, which strongly argues that L4 is a modular domain that defines MxA specificity for targeted viral structures.

*L4 influences binding of MxA to THOV NP*

We previously identified the viral NP as the target of MxA antiviral action (Zimmermann et al., 2011) and demonstrated interaction of hsMxA with THOV NP by co-immunoprecipitation (*68*). To test if the gain-of-function of the chimeric proteins is accompanied by L4-mediated binding of THOV NP, we monitored the association of NP with Flag-tagged Mx proteins in THOV-infected cells. HsMxA but not mmMx1 associated with THOV NP, as indicated by co-immunoprecipitation (**Figure 8C**). Furthermore, the mmMx1(hsL4) chimera gained the ability to bind NP (**Figure 8A and 8B**) concomitant with the gain of antiviral function. Accordingly, inserting the agmL4 into hsMxA severely reduced NP association (**Figure 9**). This indicates that L4 influences hsMxA interaction with and resultant specificity for the THOV NP. On the other hand, we did not observe rescue of NP binding by the reciprocal agmMxA(hsL4) protein (**Figure 9**). The discrepancy in NP co-immunoprecipitation by agmMxA(hsL4) versus mmMx1(hsL4) proteins may be mediated by presently unidentified residues outside of L4 that differ between the mmMx1 and agmMxA proteins, which could also have some effect (e.g., affinity or stability) on the interaction with THOV NP.

*MxA antiviral specificity for influenza A virus is mediated by L4*

We recently reported that avian influenza A viruses like contemporary H5N1 isolates are sensitive to the antiviral action of hsMxA (*62, 63*). We investigated whether the dramatic positive selection of primate MxA had any consequences on restriction of H5N1 by analyzing

the effect of hsMxA and agmMxA in Vero cells infected with avian influenza strain A/Thailand/1/04. Expression of hsMxA blocked virus replication by 10-fold (**Figure 10**); however agmMxA only had a modest effect on influenza viral replication, when compared to the inactive control hsMxA(delL4). To determine whether differences in L4 could also account for this disparity, we tested L4 chimeric proteins for antiviral activity. We found that adding the L4 region of agmMxA or the single amino acid exchange F561V to hsMxA decreased its restriction of viral replication. Conversely, adding the L4 region of hsMxA or only V561F to agmMxA increased its antiviral activity to the same level as the wildtype hsMxA protein (**Figure 10**). Taken together, these results are consistent with a generalized function for the loop L4, and F561 in particular, in MxA antiviral specificity for *orthomyxoviruses*.

**Discussion**

The diversity of viruses that MxA restricts is at odds with the observed specificity in MxA targeting of unique viral structures. Here, we leverage the antagonistic relationship between MxA and its targets using an evolution-guided approach to resolve the biological basis of MxA antiviral specificity. By analyzing signatures of recurrent positive selection characteristic of arms races between host and viral genomes, we identified a cluster of positively selected residues in loop L4 of MxA that predicts its role for antiviral specificity. Using viruses known to be susceptible to hsMxA as 'surrogates' to reveal functional consequences of ancient diversifying selection in primate MxA L4, we demonstrate that L4 is indeed a major determinant of primate MxA antiviral activity against the *orthomyxoviruses* THOV and influenza A. Given that single amino acid mutations confer MxA specificity at a site recurrently altered by positive selection, we propose that the disordered surface-exposed loop L4 mediates MxA antiviral

specificity. Although it is possible that this specificity manifests epistatically, the recurrent positive selection and demonstrated influence of L4 on the MxA - THOV NP interaction strongly implicates L4 as a direct interface of MxA target recognition.

The evolutionary signature of positive selection on MxA L4 suggests a mechanism that explains the conundrum of MxA target specificity and antiviral breadth. First, single amino acid changes in the positively selected L4 can have significant consequences for altering MxA antiviral specificity without being affected by other changes in L4 *i.e.*, F561 worked in the context of both human and AGM proteins. These experiments suggest there is relatively little epistasis between different L4 residues. We propose that other positively selected residues in L4 may determine specificity to additional viruses similar to the role that residue 561 plays as a specificity determinant for MxA against THOV. Thus, adaptation to a newly encountered pathogen could occur while not compromising pre-existing substrate specificity or antiviral range.

Second, although bearing the most striking signature of positive selection, loop L4 is not the only adaptive surface that might alter MxA specificity (**Figure 5B**). For example, great apes share identical loop L4 sequences yet differ in strength of antiviral activity suggesting that L4 is not the only domain in MxA that determines target specificity (**Figure 6A and 6B**). Moreover, the observation that hsL4 is sufficient for MxA interaction with THOV NP in the context of mmMx1 but not agmMxA also points to non-L4 determinants that influence MxA specificity. Other positively selected sites might represent distinct interaction surfaces for different classes of viral pathogens as suggested for avian Mx genes, which display marked variation in an avian-specific region in the distal N-terminus of the protein (*110*). Thus, independent specificity determinants may help MxA adapt against a wide variety of viruses. Alternatively, these sites

may function indirectly, for example by optimizing L4 positioning. Future studies will help elucidate not only the effect of L4 on MxA antiviral breadth but also whether distinct MxA surfaces are used to recognize distinct viral pathogens.

While we expect that further interrogation of diversifying sites in MxA will shed light on MxA antiviral specificity and mechanisms of restriction, it is also interesting to note that the few positively selected sites in MxA are in stark contrast to the strong purifying selection acting on the majority of the MxA protein. An explanation for the dearth of positive selection in the rest of the protein is that evolution is limited by a strictly conserved GTPase architecture. Host-host co-evolutionary interactions with cellular cofactors, such as UAP56 helicase (*111*), may further constrain MxA evolution. Therefore, flexibility in substrate recognition may have arisen in L4 to accommodate the conservation of the rest of the protein. Indeed, the transfer of hsMxA L4-mediated specificity to rodent Mx1 suggests that while the L4 function is widely conserved, divergence in L4 encodes unique antiviral specificities that reflect a history of lineage-specific arms races against ancient pathogenic viruses in primates.

Recent insights from crystallographic data also help put our findings in biochemical context. Although L4 was unresolved in the hsMxA structure, the resulting structural model situates the loop L4 facing outward from the inner surface of a ring-shaped MxA oligomer in which the G domains are at the periphery and the stalks assemble in the middle in a crisscross pattern (*49, 56*). This configuration permits extensive, repetitive contacts between the assembled MxA oligomer and its substrate. The model is particularly appealing for THOV and other negative-sense RNA viruses, because the putative viral substrate NP is repetitively bound to the viral RNA forming a large superhelical structure (*68, 112*). MxA oligomerization and cooperative target binding may also explain the large phenotypic consequences for antiviral

activity observed for single amino acid mutations in L4. Even subtle differences in the binding affinity between MxA and NP may largely determine antiviral specificity because their effect is amplified synergistically on a multimeric target like the viral nucleocapsid (*58*).

How the comparatively diminutive host innate immune repertoire withstands the diversity and rapid evolution of pathogens is a long-standing problem. Taken together, our data provide insight into the molecular nature of antiviral breadth exhibited by one effector of innate immunity, MxA, in response to pathogen diversity. Our work suggests that breadth may be achieved through the evolution of specificity domains that lie within evolutionarily dynamic and structurally 'flexible' regions of otherwise conserved antiviral proteins.

**Key for species abbreviations in alignment of MxA protein sequence**

col - *Colobus guereza* - colobus monkey
agm - *Chlorocebus pygerythrus* - African green monkey
flm - *Trachypithecus francoisi* - Francois' leaf monkey
fas - *Macaca fasicularis* - crab-eating macaque
rhe - *Macaca mulatta* - rhesus macaque
smg - *Cercocebus atys* - sooty mangabey
syl - *Macaca sylvanus* - Barbary macaque
tal - *Miopithecus talapoin* - talapoin monkey
agi - *Hylobates agilis* - agile siamang
sia - *Symphalangus syndactylus* - island siamang
wcg - *Nomascus leucogenys* - white-cheeked gibbon
gor - *Gorilla gorilla* - gorilla
hsa - *Homo sapiens* - human
bon - *Pan paniscus* - bonobo
cpz - *Pan troglodytes* - chimpanzee
ora - *Pongo pygmaeus* - orangutan
tit - *Callicebus moloch* - dusky titi
wfs - *Pithecia pithecia* - white-faced saki
how - *Alouatta sara* - Bolivian red howler monkey
mar - *Callithrix jacchus* - common marmoset
rht - *Saguinus midas* - red-handed tamarin
squ - *Saimiri sciureus* - common squirrel monkey
woo - *Lagothrix lagotricha* - common woolly monkey
owl - *Aotus trivirgatus* - three-striped night monkey

**Protein sequence for MxA from 24 primate species[###]**

```
col             MVVSEVDVVKADPAAASHPLLLNGDADVAQKSPGSVAENNLCSQYEEKVRPCIDLVDSLR 60
agm             MVVSEVDIVKADPAAASQPLLLNGDADVAQKSPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
flm             MVVSEVDIVKADPAAASRPLLLNGDADVAQKSPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
fas             MVVSEVDIVKADPAAASQPLLLNGDADVAQKSPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
rhe             MVLSEVDIVKADPAAASQPLLLNGDADVAQKSPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
smg             MVVSEVDIVKADPAAASQPLLLNGDADVAQKSPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
syl             MVVSEVDIVKADPAAASQPLLLNGDADVAQKSPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
tal             MVVSEVDIVKADPAAASQPLLLNGDADVAQKSLVSVAENNLCSQYEEKVRPCIDLIDSLR 60
agi             MVVSEVDIAKADPAAASHPLLLNGDDNVAQKNPRLVAENNLCSQYEEKVRPCIDLIDSLR 60
sia             MVVSEVDIAKADPAAASHPLLLNGDDNVAQKNPRLVAENNLCSQYEEKVRPCIDLIDSLR 60
wcg             MVLSEVDIAKADPAAASHPLLLNGDDNVAQKNPRLVAENNLCSQYEEKVRPCIDLIDSLR 60
gor             MVVSEVDIAKADPAAASHPLLLNGDANVAQKNPGLVAENNLCSQYEEKVRPCIDLIDSLR 60
hsa             MVLSEVDIAKADPAAASHPLLLNGDATVAQKNPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
bon             MVVSEVDIAKADPAAASHPLLLNGDANVAQKNPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
cpz             MVVSEVDIAKADPAAASHPLLLNGDANVAQKNPGSVAENNLCSQYEEKVRPCIDLIDSLR 60
ora             MVVSEVDIAKADPAAASHPVLLNGDANVAQKNLGSVAENNLCSQYEEKVRPCIDLIDSLR 60
tit             MVVSEVGVGKADSAAESHHLLVNGDANVAKKNPASVAENNLCSQYEEKVRPCIDLIDSLR 60
wfs             MVVSEVGVGKADPAAESCHLLVNGDANVAKKNPASVAENNLCSQYEEKVRPCIDLIDSLR 60
how             MVVSEVGVGKADPAAEPHHLLLNGDANVAKKNPALVAENNLCSQYEEKVRPCIDLIDSLR 60
mar             MVLSEVGVGKADPAVESHHQLLNGDANVAKKNPASVAENNLYSQYEEKVRPCIDLIDSLR 60
rht             MVVSEVGVGKADPAAESHHLLLNGDANVAKKNLASVAENNLYSQYEEKVRPCIDLIDSLR 60
squ             MVVSEVGVGKVDPAAESHHLLLNEGANVAKKNPALVAENNLYSQYEEKVRPCIDLIDSLR 60
woo             MVLSEVGVGKADPAAESHHLLLNGDANVAKKNPDLVAENNLCGQYEEKVRPCIDLIDSLR 60
owl             MVLSEVGVGKADPAAESHHLLLNGDANVAKKNPDLVAENNLCSQYEEKVRPCIDLIDSLR 60
                **:***.: *.*.*. .   *:* .  **:*.   ****** .************:****
```

```
col             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNKDEW 120
agm             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDEW 120
flm             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDEW 120
fas             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDEW 120
rhe             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDEW 120
smg             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDEW 120
syl             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDEW 120
tal             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDEW 120
agi             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
sia             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
wcg             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
gor             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
hsa             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
bon             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
cpz             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
ora             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEDKW 120
tit             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLMN-EEW 119
wfs             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLMNKEEW 120
how             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLMN-EEW 119
mar             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEEEW 120
rht             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEAEW 120
squ             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEEEW 120
woo             ALGVEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEEEW 120
owl             ALGAEQDLALPAIAVIGDQSSGKSSVLEALSGVALPRGSGIVTRCPLVLKLKKLVNEEEW 120
                ***.**************************************************:*  :*

col             RGKVSYQDYEIEILDASEVEKEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
agm             RGKVSYQDYEIEILDASEVEKEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
flm             RGKVSYQDYEIEILDASEVEKEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
fas             RGKVSYQDYEIEILDASEVEKEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
rhe             RGKVSYQDYEIEILDASEVEKEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
smg             RGKVSYQDYETEILDASEVEEEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
syl             RGKVSYQDYEIEILDASEVEEEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
tal             RGKVSYQDYEIEILDASEVEKEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
agi             RGKVSYQDYEIEISDASEVEKEINKAQNTIAGEGMGISHELITLEISSQDVPDLTLIDLP 180
sia             RGKVSYQDYEIEISDASEVEKEINKAQNTIAGEGMGISHELITLEISSQDVPDLTLIDLP 180
wcg             RGKVSYQDYEIEISDASEVEKEINKAQNTIAGEGMGISHELITLEISSQDVPDLTLIDLP 180
gor             RGKVSYQDYEIEISDASEVEKEINKAQNTIAGEGMGISHELITLEVSSRDVPDLTLIDLP 180
hsa             RGKVSYQDYEIEISDASEVEKEINKAQNAIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
bon             RGKVSYQDYENEISDASEVEKEINKAQNAIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
cpz             RGKVSYQDYENEISDASEVEKEINKAQNAIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
ora             RGKVSYQDYEIEISDASEVEKEINKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
tit             KGKVSYQDFEIEISDASEVEKEVNKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 179
wfs             KGKVSYQDFEIEISDASEVEKEVNKAQNTIAGEGMGISDELITLEISSRDVPDLTLIDLP 180
how             KGKVSYQDLEIEISDASEVEKEVNKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 179
mar             KGKVSYQDFEIEISDASEVEKEVNKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
rht             KGKVSYQDFEIDISNASEVEKEVNKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
squ             KGKVSYRDFETEISDASEVEREVNKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
woo             KGKVSYQDFEIEISNASEVEKEVNKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
owl             KGKVSYQDFEIEISNASEVEKEVNKAQNTIAGEGMGISHELITLEISSRDVPDLTLIDLP 180
                :*****:* * :* :*****.*:*****:*********.******:**:***********

col             GITRVAVGNQPADIGYKIKTLIRKYIQRQETINLVVVPSNVDIATTEALSMAQEVDPEGD 240
agm             GITRVAVGNQPPDIGYKIKTLIRKYIQRQETINLVVVPSNVDIATTEALSMAQEVDPEGD 240
flm             GITRVAVGNQPADIGYKIKTLIKKYIQRQETINLVVVPSNVDIATTEALSMAQEVDPEGD 240
fas             GITRVAVGNQPPDIGYKIKTLIRKYIQRQETINLVVVPSNVDIATTEALSMAQEVDPEGD 240
rhe             GITRVAVGNQPPDIGYKIKTLIRKYIQRQETINLVVVPSNVDIATTEALSMAQEVDPEGD 240
smg             GITRVAVGNQPPDIGYKIKTLIRKYIQRQETINLVVVPSNVDIAATEALSMAQEVDPEGD 240
syl             GITRVAVGNQPPDIGYKIKTLIRKYIQRQETINLVVVPSNVDIATTEALSMAQEVDPEGD 240
tal             GITRVAVGNQPPDIGYKIKTLIRKYIQRQETINLVVVPSNVDIATTEALSMAQEVDPEGD 240
```

```
agi             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
sia             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
wcg             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
gor             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
hsa             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
bon             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
cpz             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
ora             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
tit             GITRVAVGNQPADIGRKIKALIRKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 239
wfs             GITRVAVGNQPADIGRKIKALIRKYIKRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
how             GITRVAVGNQPADIGRKIKALIRKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 239
mar             GITRVAVGNQPADIGRKIKALIRKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
rht             GITRVAVGNQPADIGYKIKTLIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
squ             GITRVAVGNQPADIGRQIKALIRKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPEGD 240
woo             GITRVAVGNQPADIGRKIKALIKKYIQRQETISLAVVPSNVDIATTEALSMAQEVDPDGD 240
owl             GITRVAVGNQPADIGRKIKALIKKYIQRQETISLVVVPSNVDIATTEALSMAQEVDPDGD 240
                ***********.*** :**:**:***:*****.*.*********:************:**

col             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMVVKCRGQQEIQDQLSLSEALQREK 300
agm             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
flm             RTIGILTKPDLVDKGTEDQVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
fas             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
rhe             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
smg             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
syl             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
tal             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
agi             RTIGILTKPDLVDKGTEDKIVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
sia             RTIGILTKPDLVDKGTEDKIVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
wcg             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
gor             RTIGILTKPDLVDRGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
hsa             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
bon             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
cpz             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
ora             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALQREK 300
tit             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALEREK 299
wfs             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALEREE 300
how             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALEREK 299
mar             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALEREK 300
rht             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALEREK 300
squ             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALEREK 300
woo             RTIGILTKPDLVDNGTEDKVVDVVRNLVFHLKKGYMIVKCRGQQEIQDQLSLSEALEREK 300
owl             RTIGILTKPDLVDKGTEDKVVDVVRNLVFHLKKGYMIVKCRGRQEIQDQLSLSEALEREK 300
                *************.****::*****************:*****:*************:**:

col             IFFQDHPHFRDLLEEGKATIPCLAEKLTSELIAHICKSLPLLENQIKESHQGITEELQKY 360
agm             IFFEDHPHFRDLLEEGKATIPCLAEKLTSELIAHICKSLPLLENQIKESHQGITEELQKY 360
flm             VFFQDHPHFRDLLEEGKATVPCLAEKLTSELIAHICKSLPLLENQIKESHQGITEELQKY 360
fas             IFFEDHPHFRDLLEEGKATIPCLAEKLTSELVAHICKSLPLLENQIKESHQGITEELQKY 360
rhe             IFFEDHPHFRDLLEEGKATIPCLAEKLTSELIAHICKSLPLLENQIKESHQGITEELQKY 360
smg             IFFEDHPHFRDLLEEGKATIPCLAEKLTSELIAHICKSLPLLENQIKESRQGITEELQKY 360
syl             IFFEVHPHFRDLLEEGKATIPCLAEKLTSELIAHICKSLPLLENQIKESHQGITEELQKY 360
tal             IFFEDHPHFRDLLEEGEATIPCLAEKLTSELIAHICKSLPLLENQIKESHQGITEELQKY 360
agi             IFFEDHPYFRDLLEEGKATVPCLAEKLTSELIAHICKSLPLLESQIKESHQTITEELQKY 360
sia             IFFEDHPYFRDLLEEGKATVPCLAEKLTSELIAHICKSLPLLESQIKESHQTITEELQKY 360
wcg             VFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLEGQIKESHQTITEELQKY 360
gor             IFFQDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQRITEELQKY 360
hsa             IFFENHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKETHQRITEELQKY 360
bon             IFFQDHPYFRDLLGEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQRITEELQKY 360
cpz             IFFQDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQRITEELQKY 360
ora             IFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQRITEELQKY 360
tit             IFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHISKSLPLLENQIKESHQKITEELQKY 359
```

```
wfs             VFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPMLENQIKESHQKITEEVQKY 360
how             IFFEDHPYFRDLLEEGKATVPCLAEKLTSELIAHICKSLPLLENQIKESHQKITEELQKY 359
mar             LFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQKITEELQKY 360
rht             VFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQKISEELQKY 360
squ             VFFEGHPYFRDLLEEGRATVPCLAEKLTSELITHICKSLPRLENQIKESHQKITEELQKY 360
woo             VFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQKITEELQKY 360
owl             VFFEDHPYFRDLLEEGKATVPCLAEKLTSELITHICKSLPLLENQIKESHQKITEELQKY 360
                :**: **:***** **.**:***********::**.**** **.****::* *:**:***

col             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEEDSRLFTRLRREFHKWDIIIEN 420
agm             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEDDSRLFTRLRREFHKWGIIIEN 420
flm             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEEDSRLFTRLRHEFHKWGIIIEN 420
fas             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEDDSRLFTRLRREFHKWGIITEN 420
rhe             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEDDSRLFTRLRREFHKWGIIIEN 420
smg             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEDDSRLFTRLRREFHKWGIIIEN 420
syl             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEDDSRLFTRLRREFHKWGIIIEN 420
tal             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEDDSRLFTRLRHEFHKWGIIIEN 420
agi             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEEDIRLFTRLRHEFHKWSTIIEN 420
sia             GVDIPEDENEKMFFLIDKINAFNQDITALIQGEETVGEEDIRLFTRLRHEFHKWSTIIEN 420
wcg             GVDIPEDENEKMFFLIDKINAFNQDITALIQGGETVGEEDIRLFTRLRHEFHKWSTIIEN 420
gor             GVDIPEDENEKMFFLIDKINAFNQDITALMQGEETVGEEDIRLFTRLRHEFHKWSTIIEN 420
hsa             GVDIPEDENEKMFFLIDKINAFNQDITALMQGEETVGEEDIRLFTRLRHEFHKWSTIIEN 420
bon             GVDIPEDENEKMFFLIDKVNAFNQDISALMQGEETVGEEDIRLFTRLRHEFHKWSTIIEN 420
cpz             GVDIPEDENEKMFFLIDKVNAFNQDISALMQGEETVGEEDIRLFTRLRHEFHKWSTIIEN 420
ora             GVNVPEDENEKMFFLIDKINAFNQDITALIQGEETVGEEDIRLFTRLRHEFHKWSIIIEN 420
tit             GVDIPDDEDEKMFFLIDKINVFNQDIAAFLQGEETVGEEDIRLFTRIRNEFHKWSKVIDD 419
wfs             GVDIPEDENEKMFFLIDKINAFNQDIAAFLQGEETVGEKDTRLFTRLRNEFRQWSAVIEN 420
how             GVDIPEDENEKMFFLIDKINAFNQDIAASLQGEETVGEEDTRLFTRLRNEFRKWSAVIEN 419
mar             GVDIPEDENEKMFFLIDKINAFNQDIAAFLHGEETVGEEDVRLFTKLRNEFRKWSVVIEN 420
rht             GVDIPEDENEKMFFLIDKINAFNQDIAAFLHGEETVGEEDIRLFTKLRNEFFKWSVVIEN 420
squ             GVDIPEDENEKMFFLIDKINAFNQDIAAFLQGEETVGEEDTRLFTRLRNEFRKWSAVIEN 420
woo             GVDIPEDENEKMFFLIDKINAFNQDITAFLQGEETVREEDSRLFTKLRNEFRKWSVVIEN 420
owl             GVDIPEDENEKMFFLIDKINAFNQDITAFLQGEETVREEDSRLFTKLRNEFRKWSVVIEN 420
                **::*:**:*********:*.*****:* ::* *** *.* ****::*.** :*. : ::

col             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIIKQQIKALEEPAVNMLHTVTDMV 480
agm             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIIKQQIKALEEPAVNMLHTVTDMV 480
flm             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIIKQQIKALEEPAVNMLHTVTDMV 480
fas             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVNMLHTVTDMV 480
rhe             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVNMLHTVTDMV 480
smg             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVNMLHTVTDMV 480
syl             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVNMLHTVTDMV 480
tal             NLQEGHKITSRKMQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVNMLHTVTDMV 480
agi             NFQEGHKILSRKIQKFENQYRGRELPGFVNFRTFETIMKQQIKALEEPAVDMLHTVTDMV 480
sia             NFQEGHKILSRKIQKFENQYRGRELPGFVNFRTFETIMKQQIKALEEPAVDMLHTVTDMV 480
wcg             NFQEGHKILSRKIQKFENQYRGRELPGFVNFRTFETIVKQQIKALEEPAVDMLHTVTDMV 480
gor             NFQEGHKILSRKIQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVDMLHTVADMV 480
hsa             NFQEGHKILSRKIQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVDMLHTVTDMV 480
bon             NFQEGHKILSRKIQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVDMLHTVTDMV 480
cpz             NFQEGHKILSRKIQKFENQYRGRELPGFVNYRTFETIVEQQIKALEEPAVDMLRTVTDMV 480
ora             NFQEGHKILSRKIQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVDMLHTVTDMV 480
tit             NFQESHEIMSRKIQKFENQYRGRELPGFVNYRTFETIVKRQIKALEEPAVDMLHTVTDMV 479
wfs             NFQESHEIMSRKIEKFENQYRGRELPGFVNYKTFETIVKQQIKALEEPAVDMLRTVTDMV 480
how             NFQESHEIMSRKIQKFENQYRGRELPGFVNYKTFETIVKQQIKALEEPAVDMLHTVTDMV 479
mar             NFQESHEIMSRKIQKFENQYRGRELPGFVNYRTFETIVKQQIKALEEPAVDMLHTVTDMV 480
rht             NFQQSREIISRKIQKFENQYRGRELPGFVNYKTFETIVKQQIKALEEPAVDMLHTVTDMV 480
squ             NFQASHEIMSRKIRKFENQYRGRELPGFVNYKTFETIVKQQIKALEEPAVDMLHTVTDMV 480
woo             NFKESHEIMSRKIQKFENQYRGRELPGFVNYKTFETIVKQQIKALEEPAVDMLHTVTDMV 480
owl             NFKEGHEIMSRKIQKFENQYRGRELPGFVNYKTFETIVKQQIKALEEPAVDMLHTVTDMV 480
                *:: .::* ***:.*****************::*****:::**********:**:**:***
```

```
col             RLAFTDVSMKNFEELFNLHRTAKSKIEDIRTEQERESEKLIRLHFQMEQIVYCQDQVYRS 540
agm             RLAFTDVSMKNFEELFNLHRTAKSKIEDIRTEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
flm             RLAFTDVSMKNFEELFNLHRTAKSKIEDIRTEQERESEKLIRLHFQMEQIVYCQDQVYRS 540
fas             RLAFTDVSMKNFEELFNLHRTAKSKIEDIRTEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
rhe             RLAFTDVSMKNFEELFNLHRTAKSKIEDIRTEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
smg             RLAFTDVSMKNFEELFNLHRTAKSKIEDIRTEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
syl             RLAFTDVSMKNFEELFNLHRTAKSKIEDIRTEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
tal             RLAFADVSMKNFEELFNLQRTAKSKIEDIRTEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
agi             RLAFTDVSIKNFEEFFNLHRTAKFKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
sia             RLAFTDVSIKNFEEFFNLHRTAKFKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
wcg             RLAFTDVSIKNFEEFFNLHRTAKSKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
gor             RLAFTDVSIKNFEEFFNLHRTAKSKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDHVYRG 540
hsa             RLAFTDVSIKNFEEFFNLHRTAKSKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
bon             RLAFTDVSIKNFEEFFNLHRTTKSKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
cpz             RLAFTDVSIKNFEEFFNLHRTTKSKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
ora             RLAFTDVSIKNFEEFFNLHRTAKSKIEDIRAEQEREGEKLIRLHFQMEQIVYCQDQVYRG 540
tit             RLAFTGVSIKNFEEFFNLHKTAKSKIEDIRAEQEREGEKQIRLHFRMEQIVYCQDQVYRT 539
wfs             RLAFTNVSTKNFEEFFNLHRTAKSKIEDIRAERE REGEKQIRLHFQMEQIVYCQDKVYRS 540
how             RLAFTNVSMKNFEEFFNLHRTAKSKIEDIRAEQEREGEKQIRLHFQMEQIVYCQDQVYRN 539
mar             RLAFTDVSIKNFDEFFNLHRTAKSKIEDIRAEQEREGEKQIRLQFQMEQIVYCQDQVYRS 540
rht             RLAFTDISIKNFEEFFSLYRTAKSKIEDIRTEQEREGEKQIRLQFQMEQIVYCQDQVYRN 540
squ             RVAFTDASIKNFEEFFNLHRTAKSKIEDIRAEQERESERQIRLHFQMEQIVYCQDQVYRG 540
woo             RLAFTDVSIKNFDEFFNLHRTAKSKIEDIRAEQEREGEKQIRLHFQMEQIVYCQDQVYRS 540
owl             RLAFTDVSIKNFDEFFNLHRTAKSKIEDIRAEQEREGEKQIRLHFQMEQIVYCQDQVYRS 540
                *:**:. * ***:*:*.* :*:* ******:*:***.*: ***:*:**********:***

                         hypervariable region of L4
col             ALQKVREKELEEEKK-KKSWDVGTFQPS-STDSSMEEIFQHLMAYHQEASKRISSHIPLI 598
agm             ALQKVREKELEEEKK-KKSWDVGTFQPS-STDSSMEEIFQHLMAYHQEASKRISSHIPLI 598
flm             ALQKVREKELEEEKK-KKSWDVGTFQPS-STDSSMEEIFQHLMAYHQEASKRISSHIPLI 598
fas             ALQKVREKELEEEKK-KKSWDVGTFQPS-STESSMEEIFQHLMAYHQEASKRISSHIPLI 598
rhe             ALQKVREKELEEEKK-KKSWDIGTFQPS-STESSMEEIFQHLMAYHQEASKRISSHIPLV 598
smg             ALQKVREKELEEEKK-KKSWDVGTFQFS-STESSMEEIFQHLTAYHQEASKRISSHIPLI 598
syl             ALQKVREKELEEEKK-KKSWDVGTFQSS-STDSSMEEIFQHLMAYHQEASKRISSHIPLI 598
tal             ALQKVREKELEEEKK-KKSWDVGTFQPS-STDSSMEEIFQHLMAYHQETSKRISSHIPLI 598
agi             ALQKVREKEMEEEKK-KKPWDLGVFQSSSTTDASMEELFQHLMAYHQEASKRISSHIPLI 599
sia             ALQKVREKEMEEEKK-KKPWDLGVFQSSSTTDASMEELFQHLMAYHQEASKRISSHIPLI 599
wcg             ALQKVREKEMEEEKK-KKPWDLGVFQSSSTTDASMEELFQHLMAYHQEASKRISSHIPLI 599
gor             ALQKVREKELEEEKK-KKSWDFGAFQSSSATDSSMEEIFQHLMAYHQEASKRISSHIPLI 599
hsa             ALQKVREKELEEEKK-KKSWDFGAFQSSSATDSSMEEIFQHLMAYHQEASKRISSHIPLI 599
bon             ALQKVREKELEEEKK-KKSWDFGAFQSSSATDPSMEEIFQHLMAYHQEASKRISSHIPLI 599
cpz             ALQKVREKELEEEKK-KKSWDFGAFQSSSATDPSMEEIFQHLMAYHQEASKRISSHIPLI 599
ora             ALQKVREKELEEEKK-KKSWDFGAFQSSSATDSSMEEIFQHLMAYHQEASKRISSHIPLI 599
tit             ALQKVREKELEEEKKSKKFSFGVILEHS-SAESSMEEIFQHLMAYHQEASKRISSHIPLI 598
wfs             ALQKVREKELEEEKRSKKSCFPMIVEHS-SAESSMEEIFQHLMAYHQEASKRISSHIPLI 599
how             ALQKVREKELEEEKKSKKSCFAMIPEQS-SADSSMEEIFQHLMAYHQEASKRISSHIPLT 598
mar             ALQKVREKELEEEKRSKKSCFAMVVEPS-SSESSMEEIFQHLMAYHQECSKRISSHIPLI 599
rht             ALQKVREKELEEQKSSKKSCFAMIVEHS-SADSSMEGIFQHLMAYHQEASKRISSHIPLI 599
squ             ALQKVREKEMEEEKKSNKSCFSMIAEQF-SAESSMEEIFQHLMAYHREASKRISSHIPLI 599
woo             ALQKVRENELEEEKKSKKPCFTMIVEHF-SAESSMEEIFQHLMAYHQEASKRISSHIPLI 599
owl             ALQKVRENELEEEKKSKKSCFTMIVEHF-SAESSMEEIFQHLMAYHQEASKRISSHIPLI 599
                *******:*:**:*  :*       :   :::.*** :**** ***:* **********

col             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
agm             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
flm             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
fas             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
rhe             IQFFMLQMYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
smg             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
syl             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
tal             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
agi             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
```

```
sia             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
wcg             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
gor             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
hsa             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
bon             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
cpz             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
ora             IQFFMLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
tit             IQFFVLQTYGHQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
wfs             IQFFVLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLTRLTQARRRLAQ 659
how             IQFFVLQTYGQQLQKAMLQLLQNKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 658
mar             IQFFVLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
rht             IQFFVLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRRFLKERLARLTQARRRLAQ 659
squ             IQFFVLQTYGQKLQKAMLQLLQDKDTYSWLLKERSDTSDKRRFLKERLARLAQARRRLAQ 659
woo             IQFFVLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
owl             IQFFVLQTYGQQLQKAMLQLLQDKDTYSWLLKERSDTSDKRKFLKERLARLTQARRRLAQ 659
                ****:** **::**********:******************:******:**:********

col             FPG 661
agm             FPG 661
flm             FPG 661
fas             FPG 661
rhe             FPG 661
smg             FPG 661
syl             FPG 661
tal             FPG 661
agi             FPG 662
sia             FPG 662
wcg             FPG 662
gor             FPG 662
hsa             FPG 662
bon             FPG 662
cpz             FPG 662
ora             FPG 662
tit             FPG 661
wfs             FPG 662
how             FPG 661
mar             FPG 662
rht             FPG 662
squ             FPG 662
woo             FPG 662
owl             FPG 662
                ***
```

**Figure 11. Alignment of protein sequence for MxA from 24 primate species**

Protein alignment of 24 primate MxA orthologs, generated in ClustalW. The hypervariable regoin of loop L4 is highlighted. [###]Note that the stop codon has been removed.

# Chapter 4: Multiple adaptively evolving surfaces contribute to the antiviral breadth of MxA


## Abstract

**Broad-acting antiviral effector proteins play a central role in establishing host cell-intrinsic immunity against viral pathogens. However, the evolutionary and molecular mechanisms that underlie antiviral breadth are poorly understood. Here we provide evidence that MxA uses multiple surfaces to molecularly define target specificity against distinct viruses. These target recognition elements on the loop L4 and the previously functionally unannotated N-terminus act independently, providing a model in which adaptation on one surface can occur without compromising preexisting specificity on the other. We also provide evidence that a single evolutionary transition in human MxA resulted in an expansion of antiviral breadth against rhabdoviruses and the paramyxoviruses measles virus. These results provide an evolutionary and molecular rationale for MxA antiviral breadth, and suggest a means by which cell-intrinsic immunity can tip the balance against rapidly evolving RNA viruses.**

**Introduction**

In vertebrates, the transcriptional activation of hundreds of interferon-stimulated genes (ISGs) – the so-called "antiviral state" - is the front line of cell-intrinsic immunity against viral infection. A systematic characterization of ISG function has revealed that effectors tend to act in either a broad or targeted fashion (*14*). The type of substrate engaged by ISGs defines this difference in antiviral activity. Like pattern recognition receptors, there are several well-studied examples of ISGs that have a broad antiviral range (e.g., cGAS, TETHERIN), which manifests from targeting either ubiquitous viral substrates or essential host resources (*12, 23, 113, 114*). In contrast, target specificity precludes antiviral breadth (e.g., TRIM5α) (*115, 116*). These molecular recognition events allow for the functional classification of ISGs based on the apparent dichotomy between the type of ISG target (ubiquitous/essential versus virus-specific) and ISG antiviral range (broad or narrow) (*1*).

MxA, a Dynamin-like, large GTPase, is one such broad-acting ISG, inhibiting the replication of a wide range of RNA and DNA viruses by a poorly understood mechanism (*35*). Paradoxically, MxA restriction of viral replication requires direct interaction with a viral protein target that varies across virus families. For example, MxA interacts with the orthomyxovirus nucleoprotein (NP) (*62, 64, 68*) and the hepadnavirus core protein (*73*), which are non-homologous in structure and function. A molecular rationale for how MxA maintains antiviral breadth despite specificity in target recognition remains an outstanding question.

MxA restriction of orthomyxoviruses and bunyaviruses occurs following direct binding to the nucleoprotein, suggesting that MxA recognition of negative-sense RNA viruses may occur through a common target. However, no shared epitope has been identified to explain MxA target interaction across viral families. We previously showed that MxA engages the orthomyxovirus

Thogotovirus (THOV) NP *via* the disordered loop 4 (L4). L4 also defines antiviral specificity for the bunyavirus La Crosse virus (LACV) nucleoprotein (N) (*102, 117*). Interestingly, mutagenesis of L4 residues differentially affects MxA target recognition of the THOV NP and LACV N. These findings suggest a model in which MxA antiviral activity against negative-sense RNA viruses is mediated by L4, which targets multiple viral proteins through patches of antiviral specificity residues that are at least partially non-overlapping.

Here we build upon this model by identifying a second MxA target recognition element that is functionally independent of L4. We demonstrate that the rhabdovirus vesicular stomatitis virus (VSIV) N is also a target of MxA. However, unlike L4-mediated target recognition of THOV and LACV, we find that the MxA N-terminus ($N_{term}$) defines antiviral specificity for VSIV in an L4-independent manner. Therefore, MxA antiviral breadth is mediated by multiple surfaces that act as target recognition elements. This "separation of powers" suggests an evolutionary model that allows MxA to adapt to viral pathogens on one surface without compromising preexisting specificity on the other.

**Results and Discussion**

The nucleoprotein of negative-sense RNA viruses has been suggested to be a common target of MxA (*35*). We reasoned that an understanding of how MxA recognizes additional MxA-targeted nucleoproteins would provide new insights into the molecular basis of MxA target recognition and antiviral breadth. Owing to the large fitness cost that human (HSA)-MxA exerts on VSIV (*71, 118*), we hypothesized that a) serial passage of VSIV on HSA-MxA-expressing cells would produce MxA-resistant variants, and b) causative MxA escape mutations should be concentrated in the MxA target. The passaging scheme (**Figure 12A**) was carried out in Crandell

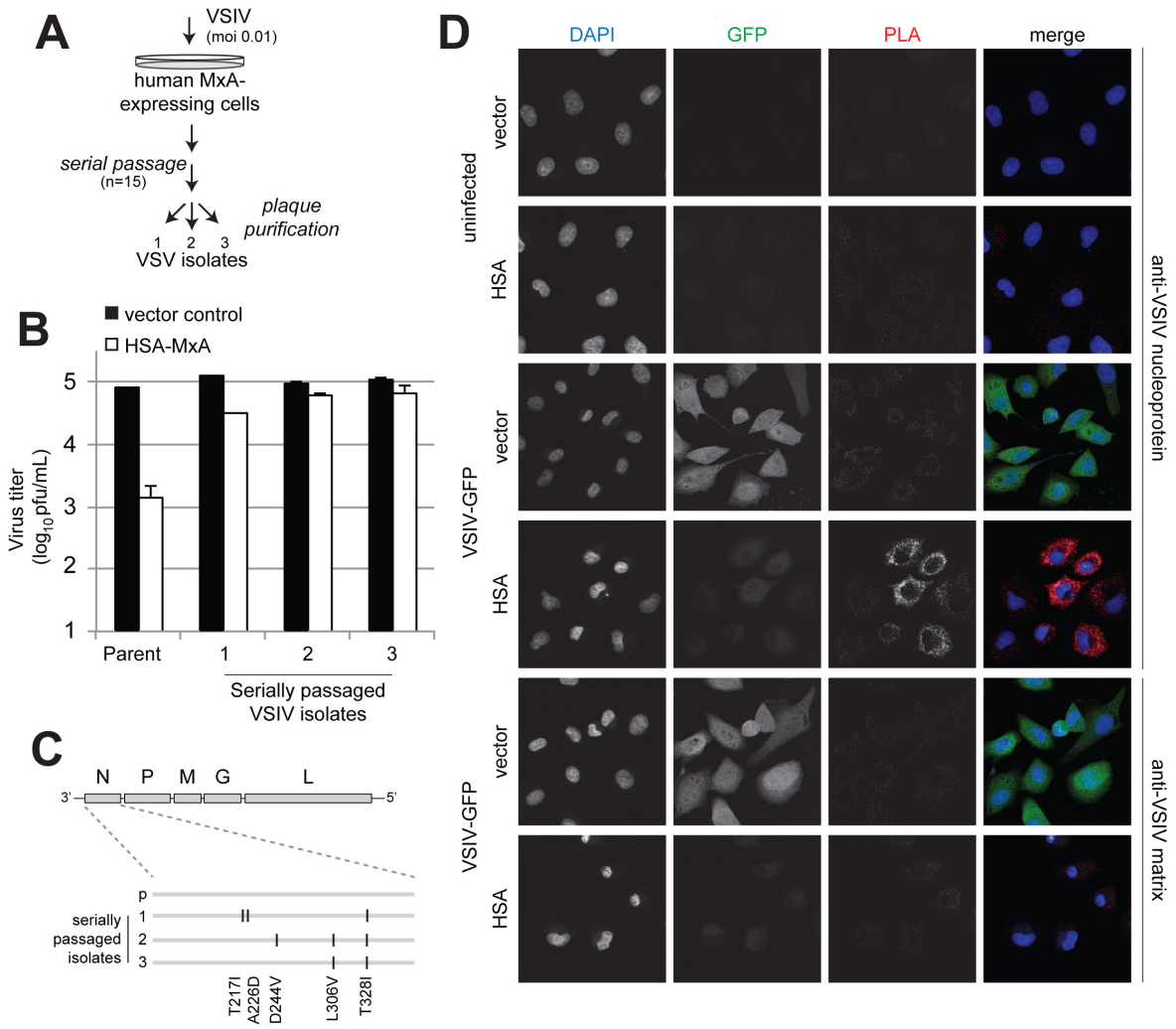



**Figure 12. Identification of the VSIV-N as the target of human MxA.**

**A.** To determine the MxA target of VSIV, virus was serially propagated on CRFK cells expressing HSA-MxA for 15 passages. Semi-clonal viral populations were isolated from three distinct plaques. **B.** Non-adapted VSIV (parent) and serially passaged VSIV isolates 1, 2 and 3 were propagated on vector control (black bars) or HSA-MxA-expressing (white bars) CRFK cells for 24h. Viral titer of each isolate was determined by plaque assay, and is presented as log10 pfu/mL. Error bars represent standard deviation from the mean. **C.** The linear, negative-sense RNA genome of VSIV is depicted, with ORFs indicated. Vertical lines denote non-synonymous mutations in the VSIV-N gene for serially passaged VSIV isolates. Black and blue lines distinguish between mixed or homogenous populations based on the frequency of mutations observable by Sanger sequencing. **D.** To establish whether HSA-MxA and VSIV-N interact, an *in situ* proximity ligation assay was conducted as described in Methods and Materials. Briefly, antibodies (Abs) raised from distinct sources (i.e., rabbit and mouse) against MxA (3xFlag) or VSIV-N were incubated on fixed and permeabilized vector control or HSA-MxA-expressing cells, that were either uninfected or infected with a GFP-reporter virus (VSIV-GFP, green) at an m.o.i. 5.0 for 5 hours. PLA probes, which contain conjugated oligonucleotides, were bound to primary Abs followed by in situ rolling circle amplification (RCA) of ligated oligonucleotides, which occurs if PLA probes are within 30-40 nm. RCA products are detected using a fluorophore-conjugated complementary probe (red). The nuclear compartment is marked by DAPI-staining (blue). Merged images that have been false-colored are shown in the far right column. The PLA procedure was also carried out using an Ab against the VSIV matrix protein.

Rees feline kidney (CRFK) cells that stably express HSA-MxA (**Figure 12B**). Following 15 passages, high titer virus was recovered from MxA-expressing cells. Three plaque-purified isolates had increased titers relative to non-passaged virus on MxA-expressing but not vector control cell lines (**Figure 12B**). Given the MxA-specific phenotype, the consensus sequence for each plaque-purified isolate was determined to identify clustered mutations. All but two non-synonymous mutations were located in the VSIV-N (**Figure 12C**). This pattern of mutations is consistent with the VSIV-N being the target of MxA. We were unable to verify an MxA:VSIV-N interaction by co-immunoprecipitation; however, the oligomeric nature of both host and viral components may hinder biochemical approaches to confirm these interactions. To circumvent these technical challenges, we attempted to visualize the MxA:VSIV-N interaction *in situ* using a proximity ligation assay. Foci consistent with staining of VSIV-induced cytoplasmic inclusions (*119*) were observed in HSA-MxA-expressing cells infected with VSIV, but not in uninfected or empty vector control cells (**Figure 12D**). The specificity of staining for an MxA:VSIV-N interaction was confirmed by using an antibody directed against the VSIV matrix protein, which did not result in PLA foci. Taken together, these data indicate that VSIV-N is the target of MxA.

To determine if L4 mediates MxA antiviral activity against VSIV, we generated a primate MxA cell line diversity panel (**Figure 13A**). The primate MxA gene has evolved rapidly under strong positive selection (*102*), consistent with a history of pathogen-driven evolution (*34*). Maximum-likelihood tests to identify positively selected codons previously revealed multiple "hotspots" of recurrent positive selection on MxA, most notably in the hypervariable region of L4 (amino acids 558-572). We therefore selected representative primates that vary at "hotspots" to leverage natural variation at positions likely to underlie phenotypic differences. Primate fibroblasts with undetectable levels of endogenous MxA were engineered to stably express MxA

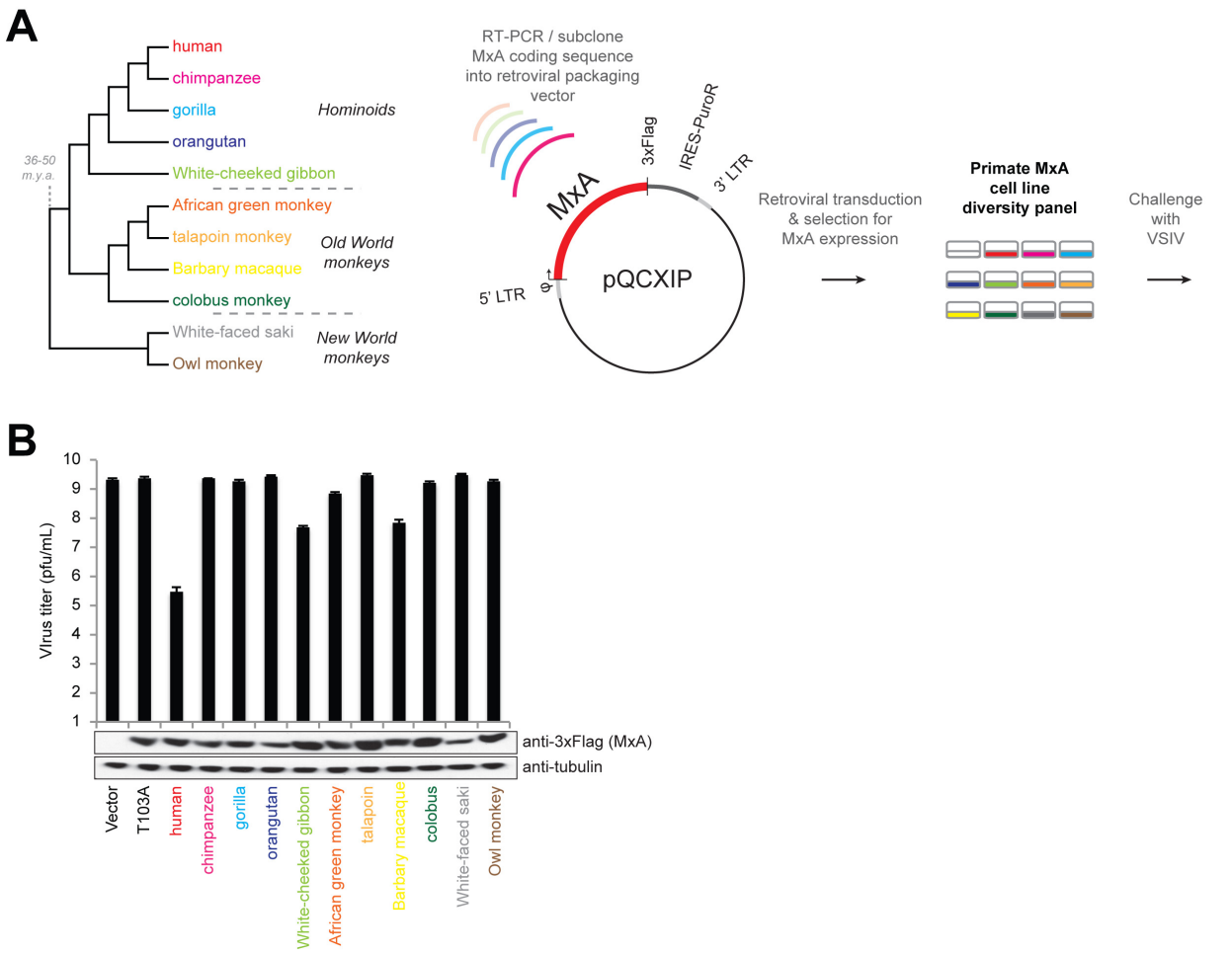



**Figure 13. Species-specific antiviral activity of primate MxA against VSIV.**

**A.** We developed a system to leverage the sequence variation across primate MxA orthologs to map determinants of antiviral activity. MxA coding sequence from selected species, which encompasses 36-50 million years of primate evolution, was subcloned into a packaging plasmid pQCXIP to allow retroviral transduction and the generation of MxA-expressing cell lines that are otherwise isogenic. The phylogenetic relationship of primate species from which MxA orthologs were derived is shown with major primate subclades indicated. **B.** *Top*, MxA-expressing cell lines were infected with VSIV at an m.o.i. 0.01. Virus-containing supernantant was collected 24 hours post-infection and viral titers (log10 pfu/mL) were determined by plaque assay. Error bars represent standard deviation between three representative experiments. *Bottom*, The expression of MxA and β-tubulin in cell lines was verified by western blot.

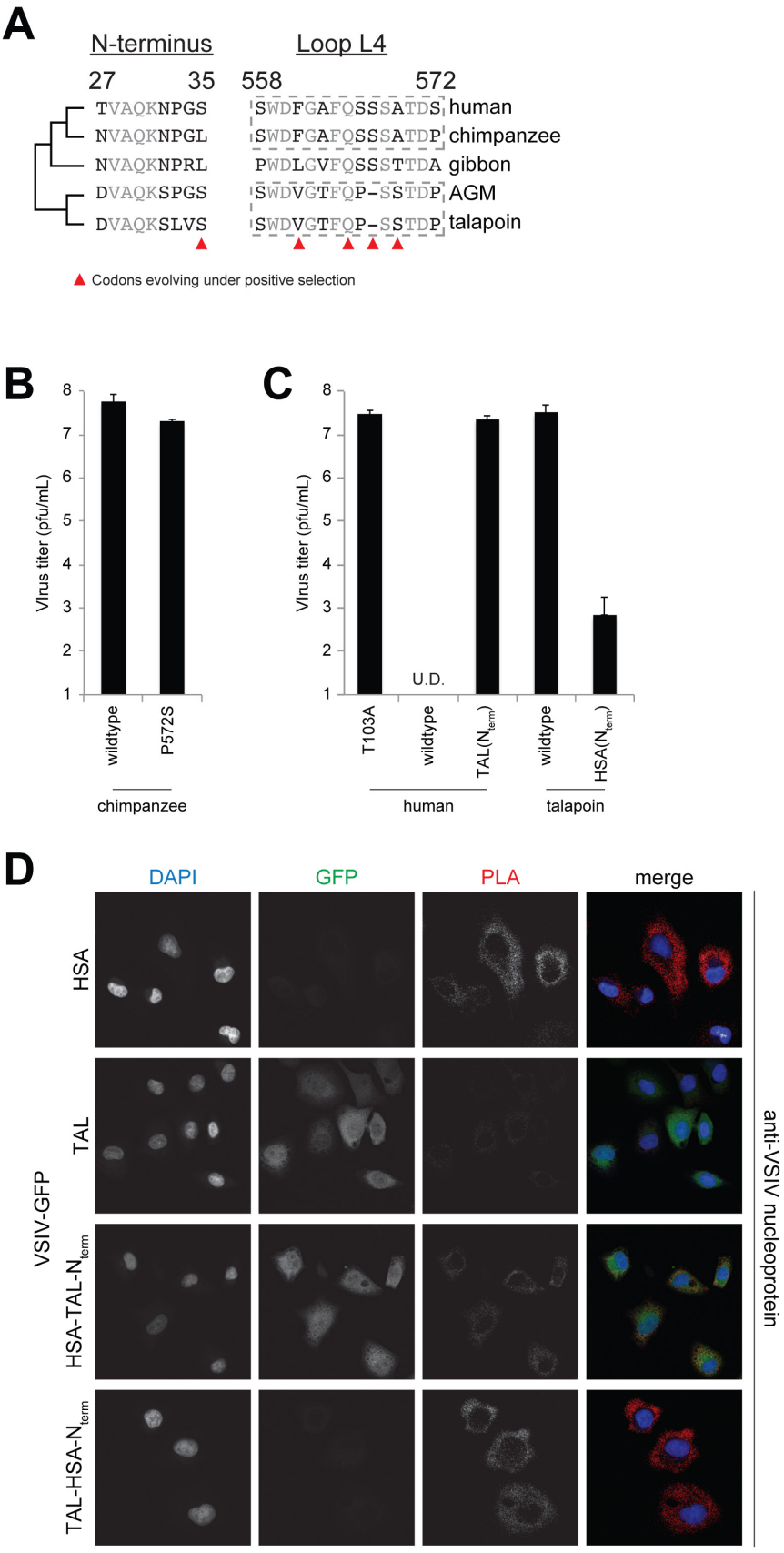

A
N-terminus
Loop L4
27
35
558
572
TVAQKNPGS
SWDFGAFQSSSATDS
human
NVAQKNPGL
SWDFGAFQSSSATDP
chimpanzee
NVAQKNPRL
PWDLGVFQSSSTTDA
gibbon
DVAQKSPGS
SWDVGTFQP-SSTDP
AGM
DVAQKSLVS
SWDVGTFQP-SSTDP
talapoin
Codons evolving under positive selection
B
Virus titer (pfu/mL)
wildtype
P572S
chimpanzee
C
Virus titer (pfu/mL)
U.D.
T103A
wildtype
TAL(Nterm)
wildtype
HSA(Nterm)
human
talapoin
D
DAPI
GFP
PLA
merge
HSA
TAL
HSA-TAL-Nterm
TAL-HSA-Nterm
VSIV-GFP
anti-VSIV nucleoprotein

**Figure 14. The MxA N-terminus is a determinant of antiviral activity against VSIV.**

**A.** A protein alignment of the $N_{term}$ and L4 regions of MxA from selected primate orthologs. Darker and lighter font indicates variable or conserved positions, respectively. The grey, dashed boxes highlight conservation between human and chimpanzee or African green monkey and talapoin MxA loop L4 sequences. Sites previously determined to have evolved under positive selection (posterior probability $PP > 0.9$) are marked by red triangles (*102*). **B-C.** MxA-expressing cell lines were infected with VSIV at an m.o.i. 0.01. Virus-containing supernantant was collected 24 hours post-infection and viral titers (log10 pfu/mL) were determined by plaque assay. Error bars represent standard deviation between three representative experiments. **D.** To establish whether the MxA $N_{term}$ targets the VSIV-N, the *in situ* proximity ligation assay for conducted in wildtype human and talapoin (HSA-MxA and TAL-MxA) or $N_{term}$ chimeric HSA-TAL($N_{term}$)-MxA or TAL-HSA($N_{term}$)-MxA cell lines as described in Methods and Materials and **Figure 13B**.

---

orthologs from 11 hominoids, old world monkeys and new world monkeys, which spans ~36-50 million years of primate evolution (**Figure 13A**). Control cell lines transduced with vector only or a catalytically inactive MxA mutant (HSA-MxA T103A) were highly permissive to VSIV infection. As expected, VSIV replication was strongly attenuated in HSA-MxA-expressing cells (**Figure 13B**) (*71, 118*). In addition, we found that MxA from white-cheeked gibbon, African green monkey (AGM) and Barbary macaque also reduced VSIV replication (**Figure 13B**), albeit to a varying extent, consistent with pathogen-driven evolution having shaped MxA antiviral specificity for rhabdoviruses throughout Catarrhini evolution.

The difference in anti-VSIV activity between human and non-human hominids was surprising given that the human, gorilla and orangutan L4 sequences are identical (**Figure 14A**). Chimpanzee (P572) differs at one position in L4 relative to human (S572), which has no effect on VSIV replication (**Figure 14B**). Similarly, the difference in the magnitude of restriction between HSA and AGM-MxA is independent of L4 variation, as cells expressing chimeric proteins with L4 swapped between HSA and AGM-MxA were equally permissive to VSIV infection relative to wildtype proteins (data not shown). Therefore, the stark difference in antiviral activity against VSIV across primate MxA orthologs is independent of L4.

We next focused on the MxA $N_{term}$ (defined as the first 54 amino acids prior to the first helix of the highly conserved bundle signaling element (BSE) domain) (*49, 54*). Positive selection of codons at the surface of antiviral proteins is highly indicative of a direct host-virus interface (*34*). We previously reported that sites 18 and 35 in the $N_{term}$ are rapidly evolving under positive selection in primates (**Figure 14A**) (*102*). Like L4, the $N_{term}$ is also predicted to be disordered and surface exposed. To map the determinants that underlie MxA antiviral activity against VSIV we made $N_{term}$ chimeras between HSA (active against VSIV) and talapoin (TAL)-MxA (inactive against VSIV). We found that cells expressing TAL-MxA bearing the HSA $N_{term}$ inhibited VSIV replication to levels indistinguishable from cells expressing the wildtype HSA protein (**Figure 14C**). In contrast, HSA-MxA with the talapoin $N_{term}$ lost antiviral specificity for VSIV (**Figure 14C**). Moreover, MxA-VSIV-N PLA-positive foci were observed in cells expressing either HSA-MxA or the TAL-HSA($N_{term}$)-MxA chimera, but not TAL-MxA or the HSA-TAL($N_{term}$)-MxA chimera (**Figure 14D**). Together, these findings indicate that the MxA $N_{term}$ is a target recognition element that mediates interaction with the VSIV-N protein.

To determine whether the MxA $N_{term}$ and L4 function independently, we evaluated the antiviral specificity of chimeric proteins with the $N_{term}$ or L4 swapped between HSA-MxA (antiviral specificity for THOV and VSIV) and TAL-MxA (inactive against THOV and VSIV). The amino acid at position 561 in L4 determines MxA antiviral specificity for THOV (*102*). Preliminarily, we find that exchange of the amino acid at position (F561, active; V561, inactive) results in a corresponding switch in antiviral specificity for THOV, but not for VSIV (data not shown). Likewise, MxA $N_{term}$ chimeras influence only antiviral specificity for VSIV, but not THOV. Therefore, the MxA $N_{term}$ and L4 function as independent target recognition elements to molecularly define antiviral specificity for VSIV and THOV, respectively.

These studies suggest a model in which MxA antiviral breadth is in part explained by utilizing two, functionally independent surfaces to target different viruses. However, given the broad antiviral range of MxA one expectation is that both L4 and $N_{term}$ molecularly define target specificity against multiple viruses. To address this possibility, we screened the primate MxA-expressing cell line panel for susceptibility to the paramyxovirus measles virus (MeV). As has been previously reported, we found that hMxA potently restricts MeV (*120, 121*). Surprisingly, no other primate MxA orthologs provided any protection against MeV infection. Characterization of the molecular basis for the unique antiviral specificity of hsaMxA against MeV is ongoing.

We propose that the structural partitioning of target recognition across multiple surfaces may mitigate the expected cost of altering pre-existing antiviral specificity as a consequence of pathogen-driven adaptation. Moreover, the ability of single amino acids to modulate MxA antiviral specificity suggests that multiple specificities may be molecularly encoded in both the $N_{term}$ and L4. For example, MxA antiviral specificity for THOV is strongly influenced by the L4 residue at 561, but this specificity is unperturbed by variation at positively selected sites at distal or even adjacent amino acids (*102, 117*). In contrast, mutagenesis of conserved residues in L4 non-specifically disrupts MxA antiviral activity (*56, 117*). Therefore, MxA antiviral breadth may ultimately be a product of non-overlapping molecular determinants of antiviral specificity, which are spread out across multiple surfaces. Indeed, the functional characterization of the MxA $N_{term}$ as a target recognition element highlights the notion that "hotspots" of positive selection demarcate critical antiviral surfaces, and suggests that other rapidly evolving residues on MxA may similarly contribute to its antiviral breadth against other RNA and DNA viruses. Our preliminary evidence suggests that a single evolutionary transition in the human lineage resulted

in the enhancement of antiviral activity against VSIV and the concurrent neo-specificity against MeV. The ability of the same adaptive mutation to expand rather than limit MxA antiviral activity may be a second evolutionary mechanism that contributes to antiviral breadth. This work highlights the primacy of MxA as a broad antiviral effector of cell-intrinsic immunity, and illuminates new evolutionary and molecular mechanisms by which host antiviral proteins can compete against the diversity of viral pathogens.

**Chapter 5: Perspectives**

Parts of this Chapter were modified from (*1*) with permission, Elsevier license number 3591720358649.

Arms race conflicts with pathogenic viruses present a seemingly untenable challenge to the host. In a single round of RNA virus replication it has been estimated that every possible point mutation is generated, thereby presenting a remarkable range of genetic variation on which natural selection can act (*122*). That, in combination with variation from recombination/reassortment and the maintenance of low frequency mutations in quasispecies genetic networks contributes to the rapid adaptation of viruses to evade or antagonize host defenses (*123, 124*). Considering that viruses are the most abundant genetic entity on Earth, the evolutionary deck would appear insurmountably stacked against the host. Yet, the oscillatory nature of Red Queen conflicts demands that the host, at some frequency, must prevail over the virus (*79*). Indeed, the signature of positive selection in host genes with antiviral function serves as an indelible signature of selection events that resulted in host survival and dominance, albeit temporary, over its viral adversaries (*28, 34*). Given that the metric of positive selection depends on recurrent events, signatures of positive selection certainly underestimate the frequency of evolutionary time in which the host is "winning."

The molecular mechanisms that contribute to the success of the host against viral infection are incompletely understood. The multifaceted nature of the host antiviral response in part serves to counterbalance the challenge presented by viral diversity. Here additional mechanisms that contribute to host resistance are considered based on inferences from the evolutionary history of primate MxA.

*Evolutionary and molecular mechanisms that contribute to antiviral breadth*

The broad antiviral activity of MxA suggests that it has been involved in multiple arms race conflicts throughout its history. Such a conflict-ridden past would be predicted to result in particularly strong evolutionary signatures of positive selection. Moreover, since MxA manifests its antiviral action via specific recognition of viral epitopes, "hotspots" of positive selection could therefore be used to directly identify at least some of the key determinants of MxA target recognition (*34*). In an analysis of MxA orthologs from simian primates, we identified three regions of MxA that have been repeatedly shaped by positive selection: the disordered loop L4, which protrudes from the stalk domain, the similarly disordered N-terminal extension and the alpha-helix 1c ($\alpha 1^c$) in the stalk domain (**Figure 5B**) (*102*).

Variation in L4 explains differences in antiviral activity among primate MxA orthologs against THOV and influenza A viruses (*102*). Human L4 grafted onto mouse Mx1 conferred both gain of antiviral activity and concomitant binding of Mx1 to the THOV NP. Intriguingly, MxA antiviral specificity for THOV is largely governed by a single amino acid F561 in L4, which has been recurrently mutated throughout primate MxA evolution. Thus, the loop L4 is a key determinant of MxA interaction with NP proteins from *orthomyxoviruses*. Similarly, positive selection in the MxA N-terminus also accounts for differences in MxA antiviral activity against VSIV between primate orthologs.

Although the large phenotypic effect of a single amino acid change in MxA seems extraordinary, other studies on the evolutionary dynamics between host antiviral genes and viruses have also uncovered occurrences of positively selected (adaptive) single residue changes with profound impacts on host-virus interactions (for example, (*84-101*)). The MxA structure, in combination with the evolutionary analysis, provides a model for how single amino acid changes

can have large functional outcomes. Molecular modeling predicts that MxA oligomers form a ring-like complex (*49, 54*). Interestingly, the loop L4 is positioned inward from the inner surface of the ring, suggesting that in its oligomeric state small changes elicited by single amino acids may act cooperatively to produce significant effects in viral target binding and consequently antiviral activity. Interestingly, another restriction factor TRIMCyp recognizes markedly different viral epitopes by interconverting a disordered surface loop between multiple conformations upon target recognition (*125*). A similar mechanism may also afford MxA the flexibility to recognize divergent targets by virtue of the disordered nature of L4. The orientation of the disordered N-terminus is less clear. In both cases, co-crystal structures with MxA bound to its viral epitope(s) will be an important advancement toward understanding the contributions of oligomerization and structural flexibility of target recognition elements for MxA target recognition.

The specificity encoded by individual amino acids suggests that a single target recognition element may be responsible for defining MxA antiviral specificity for multiple viruses. Not only does the L4 residue at 561 determine MxA antiviral specificity for THOV, but this specificity is also unperturbed by variation at positively selected sites at distal or even neighboring amino acids (*102, 117*). In contrast, mutagenesis at conserved residues in L4 markedly effect MxA target interactions (*56, 117*). Therefore, evolution has honed in precisely on positions that alter target specificity, but not at other sites in the putative target interface, which likely spans a region larger than a single residue. These observations infer that within a single target recognition element, adaptation to one virus can occur without compromising preexisting binding specificities to other targets.

By analogy, the presence of multiple target recognition elements on the same antiviral protein may also contribute to the robustness of antiviral breadth. The identification of the MxA N-terminus as a second determinant of antiviral specificity against VSIV is consistent with a model wherein other "hotspots" may coincide with distinct pathogen preferences. Moreover, the maintenance of L4-dependent antiviral activity against THOV despite variation in the N-terminus indicates that MxA targeting of these viruses are independently specified. In this way MxA target recognition is analogous to a Swiss Army knife: multi-functionality by discrete, independent surfaces with explicit tasks, which ultimately culminates in the remarkable antiviral breadth of MxA. However, the broad antiviral range of MxA makes it likely that future studies will uncover roles for both the L4 and N-terminus in targeting multiple viruses.

As with other ISGs that target specific viral substrates, this implies an adaptive cost such that selection for specificity to one virus will result in loss of some preexisting specificity. For example, variation in the NP of orthomyxoviruses determines susceptibility to MxA (*62-64*). Therefore, one prediction is that the acquisition of specificity for a non-susceptible NP variant may be concomitant with the loss of specificity to another. The importance of these trade-offs depends on the frequency of co-occurrences by viral pathogens, and may only be evident over the course of evolutionary time. In contrast to this prediction, the finding that the same adaptive mutation may expand, rather than limit, the antiviral range of MxA against two distantly related viruses (VSIV and MeV) provides an additional mechanism contributing to the maintenance of antiviral breadth. This feature resembles the evolution of "moonlighting" functions that are overwritten upon the essential functionality of proteins, which has been proposed to disallow significant variation away from a given optimum regardless of fluctuation in the selective environment. By extension, the trait of antiviral breadth may be evolutionarily hardwired

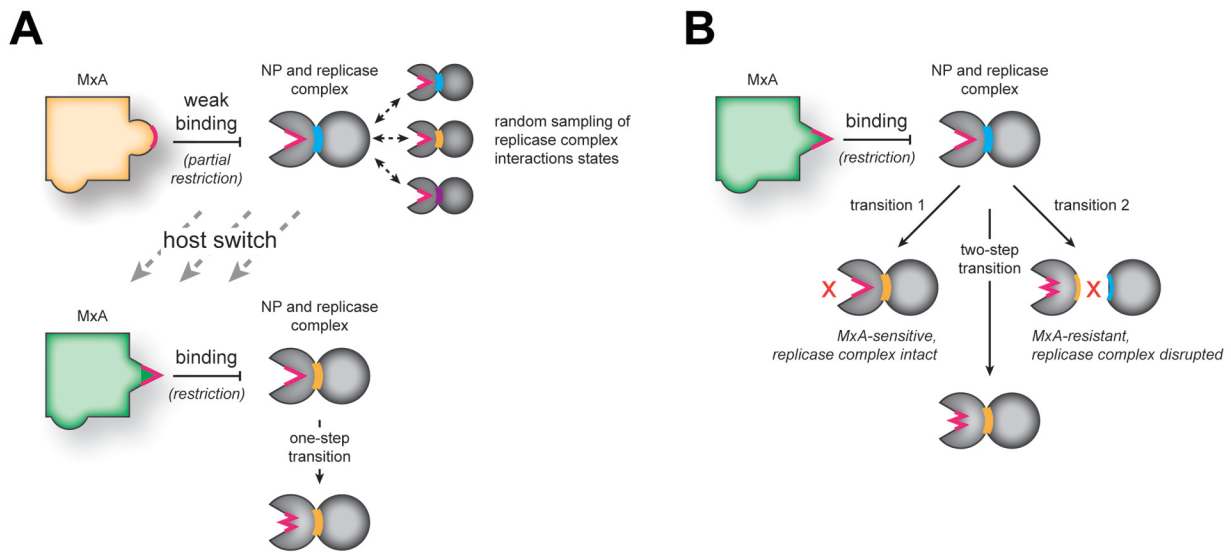



**Figure 15. The MxA-NP conflict reveals new principles governing host-virus arms races**

**A.** Virus adaptation to MxA in an intermediate host (orange) may represent an important evolutionary step in the chain of transmission for influenza and other viruses. Here, MxA binding to NP is weak such that the viral population is free to sample numerous NP/replicase 'states' (dashed lines, with sampled 'states' highlighted in blue, purple and orange). Upon transmission of the virus to a new host (green), a 'state' that permits evasion of MxA while maintaining NP/replicase interactions has been previously sampled, such that MxA can be overcome via a single evolutionary transition. **B.** MxA (green) may selectively target a viral surface on the influenza virus NP (grey) that is highly constrained for other functions. In the schematic, the evolution of NP is constrained both by its interaction with MxA (highlighted in pink) as well as the maintenance of a functional viral replicase (*126*), each representing a distinct fitness requirement. At the initial 'state' the viral population has reached local fitness peaks that satisfy both landscapes. Selective pressure from MxA forces the viral quasispecies off local optima to explore fitness valleys that represent significant barriers to sampling mutually fit solutions. One adaptive outcome for NP might be to require mutations that modify the initial NP/replicase interaction (highlighted in blue) to maintain a functional NP/replicase (highlighted in orange) but also permits evasion of MxA. The model is agnostic as to whether mutations that modify the NP/replicase interface are restricted to the NP or also occur in other components of the viral replicase complex.

into MxA, such that mutations that result in strong pathogen preference at the cost of antiviral breadth, regardless of any short-term fitness advantage, are ultimately lost in a population over evolutionary time.

*Tipping the balance: targeting functionally constrained aspects of viral replication*

The flexibility and independence of MxA target recognition notwithstanding, the antiviral utility for proteins like MxA that recognize a specific target against rapidly evolving viruses

might still be expected to be short-lived. Insight into how MxA potentially circumvents this problem comes from the recent mapping of residues on the influenza virus NP that confers Mx-resistance (*64, 65*). Unique NP segments have been independently introduced into humans by the 1918 and 2009 pandemic influenza A strains. Both NPs are uniquely resistant to human MxA relative to their avian or swine ancestor. Manz et al. identified clusters of surface exposed residues in the human-adapted NPs that conferred MxA-resistance to an avian H5N1 virus. However, the introduction of MxA-resistance mutations in the NP of avian H5N1 viruses caused significant attenuation of viral growth in the absence of MxA. Similarly, the 2009 pandemic H1N1 strain passaged in $Mx1^{-/-}$ mice acquired mutations that diminished Mx-resistance. The concomitant loss of replicative fitness with the gain of Mx-resistance suggests that several changes in NP, as well as permissive or compensating mutations in the viral polymerase, may be required for NP proteins to acquire MxA-resistance.

Selection for and maintenance of otherwise attenuating mutations in circulating human H1N1 influenza viruses strongly suggest that MxA exerts a constant selective pressure on these viruses. Indeed, 1957 and 1968 pandemic human H1N1 reassortant viruses maintained the 1918-origin NP segment despite acquisition of other avian segments, such that a novel human MxA-resistant NP has evolved only twice in the last century. These studies highlight the significance of MxA as a barrier to zoonotic transmission of avian influenza viruses. Phylogenetic analysis of archival avian, swine and human influenza sequences also indicate that NP human MxA-resistance mutations in the 2009 pandemic strain emerged in swine. Therefore, while intermediate hosts have been proposed to serve as a “melting pot” that contributes to antigenic reassortment, intermediate hosts may similarly provide an environment that enhances the likelihood that mutations that enhance human MxA-resistance arise (**Figure 15A**).

These studies suggest that MxA evasion and efficient replication may independently shape the viral fitness landscape that constrains NP evolution. An outstanding question in host-virus conflict dynamics is how the comparatively sluggish host "keeps pace" with viruses that have a mutation rate orders of magnitude beyond that of DNA-based organisms. The associated fitness cost to the virus that results from MxA evasion may provide one explanation, effectively narrowing the gap in 'evolvability' between host and viral proteins (**Figure 15B**). Therefore, although MxA viral targets may not share a common structural epitope, they may be characterized by evolutionary constraint by multifunctionality. Indeed, the identification of the VSIV-N as an MxA target is consistent with this model. Precedent for this mechanism is also exemplified by poxvirus antagonists, which subvert cellular processes by mimicking highly constrained interactions between host proteins (*127*). Given that other ISGs also target viral proteins under functional constraint, this may be a universal strategy employed by host defenses to level the battlefield in host-virus conflicts. Similar strategies have also recently been suggested to rationalize therapeutic design of novel antivirals (*128*).

These inferences into the evolutionary and molecular basis of antiviral breadth of MxA provide a framework to study the relevance and extension of these findings to other antiviral proteins.

## Chapter 6: Future Directions

Parts of this Chapter were modified from (*1*) with permission, Elsevier license number 3591720358649.

Recently, there has been remarkable growth in the field of MxA biology. These advancements have refined and expanded upon prior knowledge, largely due to the advent of new technology. For example, the collaborative cross, a genome-wide association study (GWAS) of established lab strains and wild mice, found that the quantitative trait of resistance to influenza virus could be largely explained by the Mx locus (*129*). Likewise, the structure of MxA (*49, 54*) (and MxB (*130, 131*)) has contributed a great deal to the biochemical nature of MxA antiviral activity, and more broadly to the function of Dynamin-like large GTPases (*57, 58*). This Dissertation underscores the pragmatic use of interdisciplinary approaches to advance our understanding of MxA biology. Here, I consider additional avenues of research that are uniquely poised for study by combining the disciplines of molecular evolution, virology and cell biology.

### *Part I: Insights into MxA biology from positive selection*

Positive selection in antiviral genes can be used as a signature to guide the identification of essential host-virus interfaces (*34*). In addition to the MxA L4 and N-terminus (see Chapter 3 and Chapter 4), a third "hotspot" of positive selection was also identified in $\alpha 1^c$ of the stalk. This region mediates MxA homo-oligomerization, a process that is required for MxA antiviral activity. For example, the point mutant R408D in $\alpha 1^c$ or the quadruple mutant YRGR440-443AAAA in the interacting loop 2 traps MxA into a stable dimer, resulting in loss-of-function

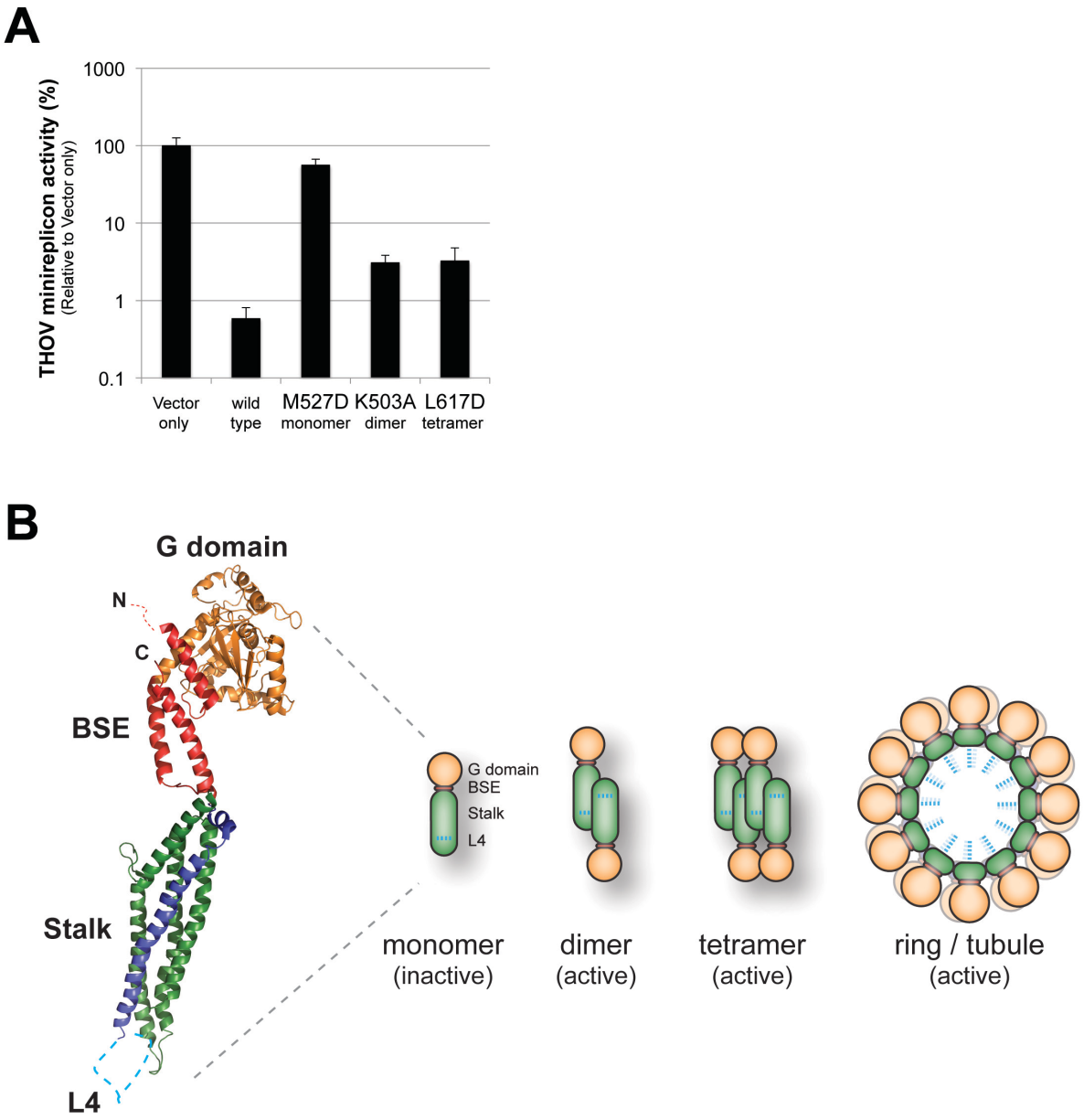



**Figure 16. MxA has multiple, antivirally active states**

**A.** THOV minireplicon activity is represented as a percentage relative to values from the empty vector control. Single point mutants have been previously shown to lock the protein in a stable monomer (M527D), dimer (K503A) or tetramer (L617D) (*49, 54*). **B.** The MxA crystal structure (3SZR) (*49*) is depicted, with colors as described for **Figure 5B**. Cartoon representations of various MxA oligomeric states is shown. Note that the $\alpha 1^{c}$ is exposed and free to interact with putative target in the dimer and tetramer.

(*54*). It is therefore surprising that a surface that underlies a conserved, essential process of MxA is also rapidly evolving. Based on the host-virus conflict model, there are two rationales for positive selection in the $\alpha1^c$: MxA is either "evading" a viral antagonist or "chasing" a viral target. To date, no direct viral antagonist of MxA has been identified. However, in theory, binding to the MxA stalk to inhibit oligomerization would be an effective strategy of viral antagonism. Consistent with this model, human MxA restriction of VSIV is markedly reduced in vaccinia virus co-infections (see also Future Directions, Part II). Alternatively, the stalk could act similarly to L4 and the N-terminus in target recognition. The $\alpha1^c$ surface is buried by BSE-stalk interactions in the context of a tetramer or higher order ring-like structure. This would necessitate $\alpha1^c$ to be free for substrate binding prior to or in the absence of oligomerization, a direct violation to the dogma of antiviral activity by GTP hydrolysis upon oligomerization. To test this hypothesis, a panel of MxA mutants were constructed that have either been shown experimentally or are predicted via modeling to block dimerization, tetramerization or the formation of higher order structures. Screening this panel for antiviral activity against the THOV minireplicon found that the MxA monomer completely lost activity. In contrast, both the locked dimer and tetramer mutants restricted THOV to a lesser but significant level relative to wildtype MxA (**Figure 16A**). This preliminary result suggests that oligomerization is not an absolute requirement of MxA antiviral activity against THOV, providing initial evidence that MxA function can be maintained in a model where the stalk is free to engage viral substrates (**Figure 16B**). Consistent with the loss-of-function in the locked monomer, $\alpha1^c$ is not involved in dimerization (*54*). These results are inconsistent with the studies of Gao et al. (*49, 54*), which indicate that only the fully oligomerization-competent MxA has antiviral activity. This discordance may be explained by the use of two different viruses (A/Vietnam/1203/04 versus

THOV). In addition, the THOV system is 10-100 times more sensitive to the effects of MxA than the analogous H5N1 system, which provides enhanced sensitivity to partial phenotypes. The existence of multiple, antivirally active MxA complexes would be predicted to expand upon the already numerous means by which inherent flexibility contributes to MxA target recognition and antiviral breadth.

*Part II: Insights into MxA biology from viruses*

The study of viruses has arguably taught us more about host biology than any other field of science (*132*). This of course also pertains to the study of MxA and its role as a molecular barrier to viral infection. Indeed, the observation that MxA-resistance mutations are acquired in the influenza virus NP upon zoonotic transmission clearly demonstrates the importance of MxA (*64*). Similarly, recent observations of our own and others' predict new avenues of research into MxA. Over the course of this Dissertation more than 30 primate MxA orthologs have been cloned and sequenced, the majority of which have been subcloned into expression constructs and stably transduced into cells. Moreover, numerous domain-swapped and point mutant chimeras have been established between orthologous species pairs. This panel provides a resource for the facile mapping of functional differences that are mediated by minimal sequence variation among closely related species. In addition, *in vitro* passaging of non-human MxA-sensitive viruses on primate MxA-expressing cell lines provides a safe means to acquire gain-of-function mutants to allow the identification of targeted viral substrates and the study mechanisms of viral evasion. In effect, the paired axes of primate and viral diversity represent a wealth of information that is both broadly applicable and delightfully esoteric. Here I provide two such examples, which highlight new avenues for MxA research driven by lessons from viruses.

*Part IIa: Poxvirus-driven evolution of primate MxA*

The vast majority of MxA-sensitive viruses are negative-sense RNA viruses. However, monkeypox virus (an orthopoxvirus) and African swine fever virus (the sole representative of *Asfarviridae*), both large, double-stranded DNA viruses that replicate in the cytoplasm, have been reported to be restricted by human MxA (*75, 133*). The molecular basis for MxA target recognition and antiviral activity of DNA viruses is unknown. We rationalized that an understanding of how MxA restricts DNA viruses may shed light on a unifying principle that underlies MxA antiviral activity. To address this question, we first attempted to replicate the ability of human MxA to restrict orthopoxvirus replication by using the prototypical orthopoxvirus vaccinia virus (VACV). Surprisingly, cell lines stably transduced with human MxA, the T103A catalytic mutant or a vector-only control were equally permissive to VACV infection (**Figure 17A**). However, multiple non-human primate MxA orthologs potently restricted VACV (**Figure 17A**). These results indicate that VACV is sensitive to variation in primate MxA, and suggest that anti-orthopoxvirus activity of primate MxA orthologs (at least human) may be sensitive to variation between VACV and monkeypox virus. This result also highlights the utility of screening otherwise isogenic primate MxA-expressing cell lines for susceptibility to viruses, not only to reveal functional consequences primate diversity, but to identify MxA-sensitive viruses that are specifically resistant to human MxA. Interestingly, VACV represents the only known example of a virus that is restricted by a NWM MxA. Moreover, primate MxA restriction of VACV is variable across and within clades, indicating that anti-orthopoxvirus activity was likely an ancestral antiviral specificity that has been recurrently

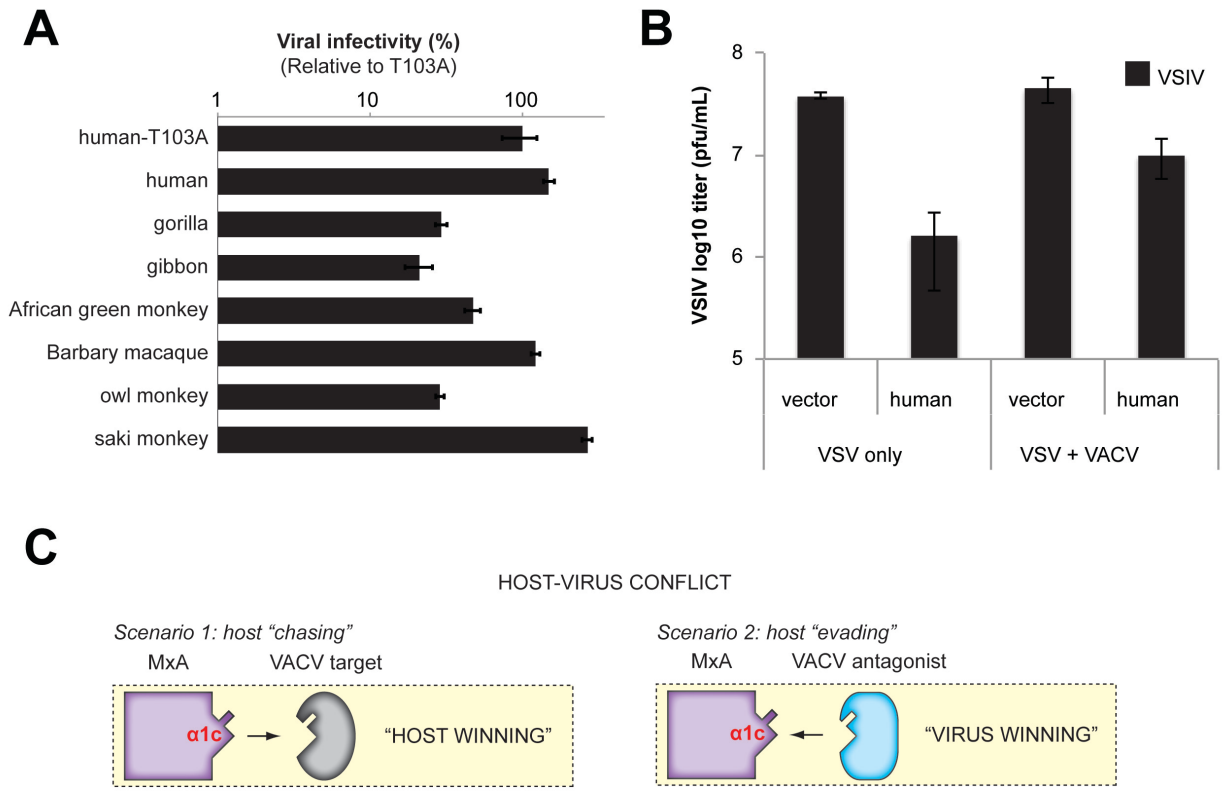



**Figure 17. Primate MxA restricts VACV replication**

**A.** Primate MxA orthologs differentially restrict VACV replication. VACV replication on MxA-expressing BSC40 cells was assessed by plaque assay on permissive cells. Viral infectivity relative to human MxA T103A is shown on a log10 scale. This figure was generated from experiments conducted by Daniel Ramirez. **B.** VACV partially rescues VSIV replication in an MxA-dependent manner. Vector or human MxA-expressing CRFK cells were infected by VSIV only or VSIV co-infected with VACV. VACV co-infections enhanced VSIV replication in MxA-expressing but not vector control cells. **C.** MxA-VACV interaction models, wherein the MxA $\alpha1^c$ is either targeting or targeted by a VACV-encoded protein. In both scenarios, $\alpha1^c$ would be predicted to evolve rapidly to establish or evade recognition.

lost and regained during primate evolution. Indeed, orthopoxviruses have been shown to mediate pathogen-driven evolution of other primate antiviral genes (*94*). The sensitivity of VACV to non-human primate MxA provides an opportunity to select for viruses that acquire mutations that rescue sensitivity to MxA in cell line passaging experiments. As previously noted, VACV co-infection rescues VSIV replication on MxA-expressing cells (**Figure 17B**). This result is consistent with VACV antagonism of human MxA, and may explain the lack of anti-VACV

activity of other primate MxA orthologs. Conducting the same experiment in cells expressing an MxA ortholog with dual anti-VACV anti-VSIV activity is an important control to verify specificity. Additionally, VACV antagonism of human MxA should similarly rescue other human MxA-sensitive viruses. To date there have been no reported examples of direct viral antagonism of MxA, a finding that would underscore the importance of MxA as a potent barrier to orthopoxvirus infection. The co-infection study may also be explained by higher expression of the VACV target or greater affinity in MxA binding, wherein the MxA-VSIV interaction would be outcompeted. Regardless, mapping MxA antiviral specificity for VACV (and/or VACV antagonism) will be illuminating. Given that there are three "hotspots" on MxA, the molecular determinants for VACV restriction are predicted to map to $\alpha 1^c$, or to either the MxA L4 or the N-terminus (**Figure 17C**). The former would reveal a novel function for $\alpha 1^c$, and the latter would provide a means by which to address how a single target recognition element on MxA encodes antiviral specificity for two distinct viruses.

*Part IIb: Rapid evolution of MxA cellular localization: a phenotype driven by flaviviruses?*

To date the strongest evidence for MxA as a single gene determinant to viral infection comes from experiments in transgenic mice (*48, 66*), in which otherwise lethal influenza virus infections are protected by expression of human MxA. However, given that only hominoids show any activity against orthomyxoviruses (*102*), the acquisition of anti-influenza virus activity is likely a more recent event in primate evolution.

The family *flaviviridae* are arthropod-borne viruses that include important human pathogens such as Dengue (DENV), West Nile (WNV) and yellow fever (YFV) viruses. Virus genetic diversity and phylogenic analyses of human and sylvatic isolates indicate the

establishment of sustained human infectious cycles by cross-species transmission from a non-human primate (NHP) reservoir, with clear evidence of an African origin (*134*).

Molecular clock estimates suggest that YFV originated in NHPs less than 100,000 years ago (*135*). However, our paleovirological studies into hepacivirus-driven evolution of the primate mitochondrial antiviral signaling protein (MAVS, or VISA, IPS1) indicate that selective pressure from hepacivirus-like ancestors dates back on the order of millions of years (*86*). This study not only dramatically alters time estimates for flavivirus evolution in primates, but also provides strong evidence that flaviviruses have been important selective pressure on primate innate immunity genes. This work and other studies highlight the strength of phylogenetic inquiry of host antiviral genes – as opposed to estimates that rely on extant virus genetic diversity – to infer the existence and age of paleoviruses (*136*). Given that the origin and subsequent radiation of other flaviviruses is largely enigmatic, leveraging the evolution of primate antiviral genes is proposed to shed light not only on the molecular determinants of flavivirus host range and cross-species transmission, but also flavivirus origins and evolution. Proof-of-principle for MxA restriction of flaviviruses has been demonstrated by artificially expressing MxA in the context of a WNV replicon, thereby bypassing MxA exclusion from otherwise protective virus-induced membrane structures (*137*). Similarly, artificial targeting of MxA to the endoplasmic reticulum (ER) resulted in robust attenuation of WNV infection (*138*). Pig Mx has also been reported to restrict the flavivirus Japanese encephalitis virus (*139*), suggesting that an evolutionarily conserved anti-flavivirus activity for MxA. These studies in concert with the pervasiveness of MxA antiviral activity against RNA viruses, and evidence for ancient and ongoing arms races between flaviviruses and primates substantiate the hypothesis that flaviviruses have been major drivers of selection on primate MxA gene evolution.

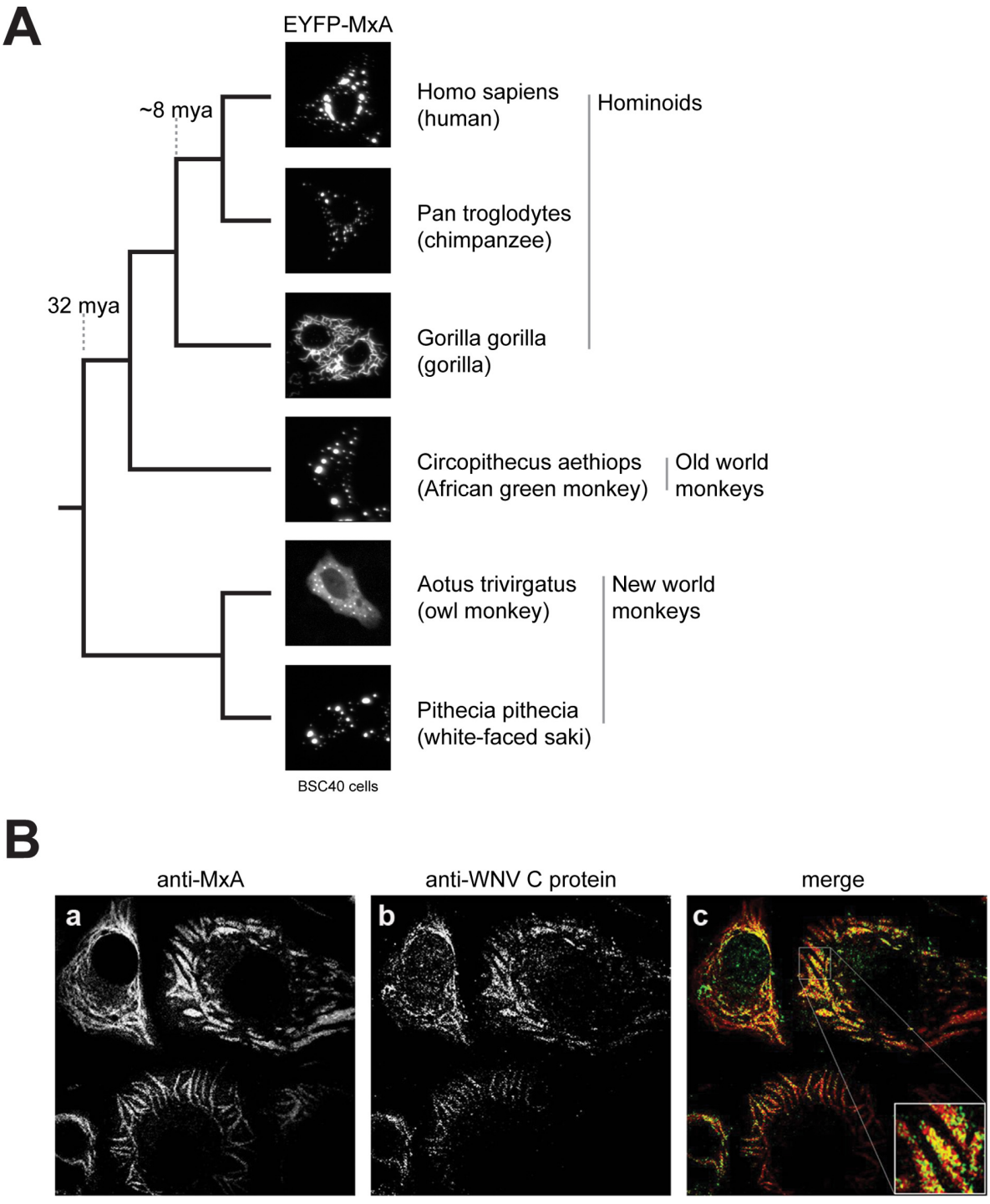



**Figure 18. Localization is a rapidly evolving phenotype of primate MxA**

**A.** Visualization of transiently expressed N-terminally tagged YFP-MxA from representative primate orthologs. Localization of gorilla and owl monkey MxA is aberrant relative to canonical foci-forming patterns observed by human MxA. **B.** MxA (a) expressed from a WNV KUNrepMxA replicon and WNV-KUN-C (b) protein co-localize (c) in tubular structures in the cytoplasm of the WNV packaging cell line BHKtet-KUNCprME. **Figure 18B** was modified from (*138*) with permission, Elsevier license number 3591720965857.

Interestingly, Gillespie et al. have reported that WNV replicon-expressed MxA leads to profound targeting of MxA to reticular membranes (**Figure 18B**) (*137, 138*). This is in stark contrast to the canonical endoplasmic reticulum-associated foci observed by MxA cytology studies in other cellular contexts (*140*). A survey of ectopic expression of primate MxA in BSC40 cells unexpectedly revealed that the localization of MxA has been dynamic throughout primate evolution (**Figure 18A**). Consistent with previous reports on human MxA, representative OWM orthologs formed ER-associated foci. In contrast, gorilla MxA formed reticular structures reminiscent of those reported by Gillespie et al. Preliminary characterization of NWM MxA orthologs has revealed additional cytological variation. The effect of artificial targeting of MxA to distinct organelle membranes on antiviral activity against flaviviruses and other viruses described herein will be paramount to discerning whether observed differences in the cellular distribution of MxA is biologically relevant.

One hypothesis to explain the difference in primate MxA localization based on the regulation of localization of guanylate-binding proteins (GBPs), another family of Dynamin-like large GTPases. GBP targeting to pathogen-associated vacuoles is dependent on a missing "self" mark deposited by IRGM, another class of interferon-stimulated GTPases (*141*). An attractive model to explain differences in MxA cytology is pathogen-driven evolution of "self" mark variation. Consistent with this model, reticular-like staining patterns of other primate MxA orthologs have also been observed in stable cell lines produced in cat cells. GBP targeting also requires a "non-self" mark, analogous to MxA targeting via L4 and the N-terminus. Therefore, in the absence of a "non-self" mark, it is unclear whether aberrant targeting would be deleterious.

Alternatively, genetic innovation in Mx proteins may also extend beyond positive selection at viral target interaction surfaces. For instance, rodent Mx paralogs have

subfunctionalized wherein Mx1 localizes to the nucleus while the recently diverged Mx2 resides in the cytoplasm. These differences in cellular localization have led to the evolution of specificity toward viruses that replicate in the respective cellular niche of each paralog (*35, 142*). In this way, rodents have split the burden of antiviral breadth by subcellular compartmentalization of two active antiviral Mx proteins. Therefore, modification of MxA localization and membrane targeting may represent an evolutionary strategy to optimize the likelihood of contacts with viral targets. Akin to evolutionary insights into MxA target recognition, a detailed cytological survey of MxA orthologs may uncover more widespread modification of MxA localization. The capacity for rapid evolution at molecular interfaces is clearly one strategy to overcome the challenge of viral diversity. Likewise, mapping the molecular basis of MxA localization will reveal whether positive selection has acted to alter the cellular distribution of primate MxA, which would represent a novel outcome of adaptive evolution in an antiviral gene. As such, a combination of traits primed for adaptation, such as subcellular localization, may ultimately explain the antiviral breadth of MxA. The confluence of these strategies is envisioned to provide a powerful model for how MxA antiviral breadth varies across cytological space and evolutionary time.

**Appendix 1: Signatures of positive selection predict pathogen-driven evolution and antiviral function of MxB**

An understanding of how ancient, pathogenic viruses (or paleoviruses) have shaped the evolution of the host antiviral repertoire has provided key insights into host-virus interactions. Such evolutionary inferences have been particularly important in describing how variation at critical molecular interfaces contribute to cross-species transmission and viral host range (*143*). The antagonistic nature of host-virus interactions drives molecular innovations by both parties in an attempt to establish or evade recognition. Such arms races have recurrently played out over the course of primate evolution. Surfaces on antiviral proteins that target (or are targeted by) viruses have rapidly evolved in a recurrent fashion, as generally adaptive mutations converge to alter the host-virus interface similarly to achieve the greatest phenotypic advantage. Signatures of positive (i.e., adaptive or Darwinian) selection can be identified by calculating rates of non-synonymous relative to synonymous mutations in an alignment of orthologous genes. Numerous studies have used such analyses to identify critical surfaces on proteins with a known role in immunity against viruses (for example, (*84-101*)). This Appendix provides proof-of-principle that positive selection analyses can also be used as a tool to implicate genes with putative antiviral function.

*The curious case of MxB: searching for antiviral activity in an interferon-induced GTPase*

Most mammalian genomes contain two tandemly arrayed Mx genes (Mx1 and Mx2), which encode MxA and MxB, respectively. MxA is an interferon-inducible Dynamin-like large GTPase, and is characterized by its remarkably broad antiviral activity. In contrast, the function

of MxB remained enigmatic until recently (*45-47*). MxA and MxB share 60% amino acid identity, which is attributed to the conservation of the GTPase architecture. Both paralogs have an N-terminal extension relative to the ancestral Dynamin. The MxB N-terminus ($MxB_N$) is distinguished by a nuclear localization signal (NLS) that underlies the prominent difference between MxA and MxB subcellular localization. MxA forms canonical foci on ER-associated membranes (*140*), although notably has been reported to form cytoplasmic tubular structures in the presence of infection or in a cell type specific fashion (*137, 138*) (also see Future Directions). In contrast, MxB is positioned on the outer surface of nuclear pores (*144*), which has been suggested to play a role in MxB-mediated regulation of cytoplasmic-nuclear trafficking. The difference in cytology notwithstanding, both MxA and MxB share many features that envisage an antiviral function for MxB. The most glaring predictive characteristic is its strong induction by type I interferons. The Mx2 gene locus contains its own interferon-stimulated response element, such that the MxA promoter likely does not drive MxB expression (*42*). Consistent with this assumption, transcriptional profiling of MxA and MxB transcripts across tissues is similar but non-overlapping (*145*). Yet decades of research using MxB as a negative control in the assessment of MxA antiviral activity against multitudes of viruses cast doubt on the putative role of MxB in cell intrinsic immunity.

*Positive selection in MxB suggests a history of host-virus conflicts during primate evolution*

We previously showed that MxA has undergone episodic positive selection during the radiation of primates, consistent with a history of pathogen-driven evolution (*102*). To determine whether there is evidence for pathogen-driven evolution of MxB, orthologs from 12 hominoids, Old World monkeys and New World monkeys were cloned and sequenced (for a total of 22

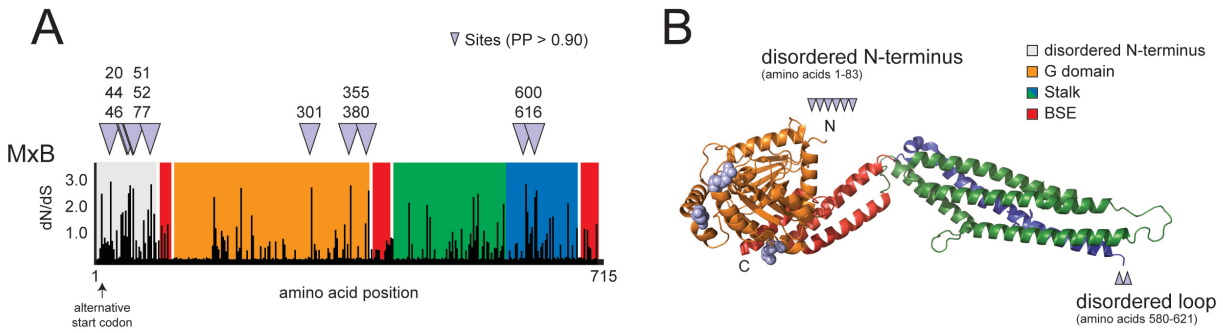



**Figure 19. Rapid evolution in MxB is concentrated in the disordered N-terminus and loop L4**

**A.** dN/dS for each codon as determined by Bayes empirical Bayes (BEB) implemented using PAML model M8 on alignments of the full length gene and the $MxB_N$ and loop L4. Sites for which posterior probability (PP) of dN/dS > 1 exceeds 0.90 are indicated by a triangle. Protein domains are overlayed with colors corresponding to the crystal structure in Figure A1B. An arrow marks the alternative start codon at position 23. **B.** The MxB crystal structure (PDB 4WHJ) (*130*) is shown, with sphere representations of rapidly evolving sites in the G domain. Enrichment of positively selected sites in the disordered N-terminus and loop L4 are indicated by triangles, as in Figure A1A. Illustrative rendering of the MxB crystal structure was conducted using PyMol (*50*).

sequences when combined with publicly available genomes). Maximum likelihood-based methods were used to assess the rate of non-synonymous (dN) relative to synonymous (dS) changes in an alignment of primate MxB coding sequences. This analysis revealed evidence for strong positive selection ($P < 0.001$) having acted on MxB during primate evolution (**Table 2**).

These analyses also revealed multiple rapidly evolving sites on MxB surfaces (**Figure 19A and 19B**). Clustering of such sites, or "hotspots" of positive selection on antiviral proteins often occur at positions in direct contact with viral proteins (*34*). For example, a clustering of positively selected sites guided the functional characterization of the MxA loop L4 as a target recognition element. Interestingly, nine out of 11 rapidly evolving sites (BEB poster probability PP > 0.90) were located in the disordered $MxB_N$ (amino acids 1-83) and the G domain. Indeed, when analyzed alone the $MxB_N$ shows a signature of positive selection ($P < 0.001$). In contrast, statistical support of positive selection in the rest of the protein is weak (P = 0.02). The notable exception to this is a disordered loop (hereafter referred to as $MxB_{L4}$) that sits in an analogous position to the MxA loop L4. Two sites in $MxB_{L4}$ are evolving under positive selection in

**Table 2. PAML analysis for positive selection in primate Mx2 strong**

| MxB gene | Region analyzed | 2 (ln λ)[a] | *P* value[b] |
|---|---|---|---|
| Full length | 1-715 | 20.14 | < 0.001 |
| N only | 1-83 | 15.97 | < 0.001 |
| Δ N | 84-715 | 7.29 | 0.02 |
| L4 only | 580-621 | 2.27 | 0.32 |
| Δ L4 | 1-579::622-715 | 15.93 | < 0.001 |
| Δ N, Δ L4 | 84-579::622-715 | 3.62 | 0.16 |

[a] Results of ML tests for positive selection, showing twice the log difference between evolutionary models that allow (M8) or disallow (M7) positive selection.
[b] *P* values describe whether either model was a significantly better fit to the data.

primates. Masking these "hotspots" from the analysis completely abolishes signatures of positive selection in MxB. Therefore, primate MxB has undergone significant adaptive remodeling in a fashion that is both similar ($MxB_{L4}$) and distinct ($MxB_N$) from the rapid evolution of its paralog MxA.

*Experimental validation for evolutionary predictions of MxB function*

The signatures of positive selection in MxB, in concert with its regulation by IFN and relatedness to the antiviral MxA, strongly predict a central role for MxB in host-virus interactions. This evolutionary forecasting came to fruition with three concurrent studies demonstrating MxB restriction of simian lentiviruses, including human immunodeficiency virus type 1 (HIV-1) (*45-47*). The field has approached these results with some caution. The effect of human MxB on HIV-1 replication is modest (5-10-fold) relative to inhibition by *bona fide* host restriction factors (e.g. Apobec3G, Tetherin). In light of the rapid evolution of MxB in primates, and considerable evidence for lentivirus-driven evolution of primate host restriction factors (*28*), MxB function is likely species-specific and highly impacted by variation across simian lentiviruses. Indeed, the arms race model predicts a subtle effect for human MxB against the human-adapted HIV-1. The survey of primate diversity in this study provides a means to test the

hypothesis that primate MxB orthologs are more highly active against non-autologous virus. Thus, studying a broad range of primate MxB orthologs is useful both from evolutionary and virological perspectives.

The clustered pattern of rapidly evolving sites implicates the $MxB_N$ as a critical interface between MxB and counter-evolving pathogens. Busnadiego et al. found that variation at a single amino acid could reverse the difference in HIV-1 restriction between rhesus macaque and African green monkey MxB (G37 in rhesus macaque, R37 in African green monkey) (*99*). The authors, who also conducted an analysis for positive selection[a], found evidence for adaptive evolution at this site. This is surprising given that in our dataset only three independent changes have occurred at the site during primate evolution. In fact, with the exception of African green monkeys (G37R) position 37 is completely conserved in hominoids and OWMs (BEB PP = 0.50). Regardless, sites proximal to position 37 have clearly undergone recurrent, adaptive evolution (Figure 1). Moreover, Goujon et al. have shown that grafting the $MxB_N$ onto MxA is sufficient to transfer anti-HIV-1 antiviral specificity (*148*), indicating that like the MxA loop L4, $MxB_N$ acts as a modular target recognition element.

The viral component of this arms race appears to be the lentiviral capsid. For example, *in vitro* passaging of HIV-1 under restrictive (i.e., MxB overexpression) conditions selected for a single mutation in capsid that enhances replication in MxB-expressing cells (*45*). A survey of well-characterized HIV-1 capsid mutants, the majority of which are directly adjacent to the mutation identified in the *in* vitro passaging experiment, revealed a spectrum of replicative

---

[a] The analysis herein is almost completely non-overlapping with the Busnadiego et al. study (*99*). This is likely due to power differences between datasets. For example, Busnadiego et al. found no significant difference between models M1vsM2 and M7vsM8. Instead, authors compared model M0 (which assumes a single ω for all sites, which is "probably wrong for all functional proteins" (*146*)) to model M3. The M0-M3 comparison "is a test of variability in the ω ratio among sites and does not constitute a rigorous test of positive selection" (*147*)).

fitness in an MxB-dependent manner (*46, 47*). Over-expressed or recombinant human MxB both interact with assembled HIV-1 capsid (*130, 131*). MxB-capsid binding correlates with antiviral activity, and is strongly influenced by but not dependent on the $MxB_N$, as a mutant lacking the entire N-terminus demonstrates binding over background. One intriguing possibility is that the $MxB_{L4}$ also directs MxB-capsid interaction. Therefore, the rapid evolution in MxB likely reflects recurrent selection for recognition of variant HIV-like capsids during primate evolution.

*Unique evolutionary histories reflect differences in structure and function between primate Mx orthologs*

Another striking feature of positive selection in MxB is the enrichment of sites in the G domain. The function of all Dynamin-like large GTPases (e.g., classical Dynamins, Atlastin, Mitofusin, GBPs, MxA) is strictly dependent on nucleotide binding and hydrolysis. Strikingly, classical mutations known to disrupt GTP binding (K131A) or enzymatic activity (T151A) have no effect on MxB restriction of HIV-1 (*46, 47*). It has been previously suggested that the force produced by GTP hydrolysis disrupts MxA-bound viral targets from critical interactions with other viral components (*51, 60*). Rapid evolution of the MxB G domain in light of the apparent decoupling of MxB antiviral activity from GTP hydrolysis raises the possibility that pathogen-driven evolution has resulted in the repurposing of the MxB G domain for target recognition or some other associated functionality. The conserved core typical of all GTP-binding proteins (G1 or P-loop, G2/switch1, G3/switch 2 and G4) is present in MxB; however, this may be a vestigial feature owing to its genetic ancestry. The decoupling of enzymatic and antiviral function in MxB is concomitant with the absence of higher-order oligomerization being required for antiviral activity. MxA restriction of influenza virus replication is strictly dependent on GTPase activity,

which is activated *in trans* through G domain - BSE interactions, following oligomerization into the ring-like antiviral complex (*49, 54*). Fribourgh et al. report that restriction of HIV-1 is not impeded by mutations in interface 1 of the stalk that block oligomerization (*130*). Instead, MxB forms antivirally active dimers, which would explain the GTP-independent function. It should be noted that a transition away from GTP-dependent functionality is not mutually exclusive with antiviral specificity for HIV-1, as the $MxB_N$ endows MxA with antiviral specificity for HIV-1 (*148*). We have previously identified positive selection in the MxA stalk adjacent to critical contacts that mediate oligomerization (*102*) (see also Future Directions). The ability of the non-enzymatic MxB dimer to effect antiviral activity rationalizes a model by which multiple, antiviral MxA complexes (dimer, tetramer, ring) mitigate antiviral breadth by expanding the available surfaces through which MxA can target viral substrates. Under this model, MxB may too have yet-to-be defined antiviral activity against a broad spectrum of viruses (personal communication, Sonja Best).

Taken together, these studies argue strongly that MxB is an important restriction factor for HIV-1, and may act as a molecular barrier in extant cross-species transmission events of simian lentiviruses among primates. The utility of positive selection analyses to not only pinpoint amino acid variation at critical host pathogen interfaces, but as an orthogonal approach to predict genes that function in host-virus interactions, should be widely applicable for prioritizing gene lists from screens aimed at identifying novel antiviral function.

**Appendix 2: Pervasive pathogen-driven evolution highlights the essential role of guanylate-binding proteins in primate cell-intrinsic immunity**

GTPase function is diverse and pervasive throughout biology. In particular, classical Dynamins and Dynamin-related proteins (DRPs, e.g., mitofusins, atlastins) have essential roles in vesicle scission and organelle-associated membrane remodeling (reviewed in (*53*)). While cellular "housekeeping" functions of these large GTPases are well established, the importance and extent of DRPs as effectors of the cell-intrinsic response to microbial infection has only recently been appreciated. In humans and mice there are at least 43 members of the interferon (IFN)-inducible GTPase superfamily (*149, 150*). Remarkably, this comprises greater than 20% of all proteins induced by IFNγ (*151*). Chief among these are the prototypical myxovirus resistance (Mx) proteins and the immunity-related GTPases (IRGs), which represent the most extensively studied GTPases involved in cell-intrinsic immunity (*152*). The GTPase architecture of Mx and IRG proteins is endowed through common ancestry with Dynamin. In a striking example of evolutionary convergence, another class of large GTPases, the guanylate binding proteins (GBPs), independently arose from Atlastins, and has been coopted for function in anti-pathogen defense (**Figure 20A**).

*Marked expansion and contraction of GBP genes in humans and mice: gene orthology versus functional analogy*

Genetic and functional evidence has recently demonstrated the ability of murine GBPs (mGbps) to confer cell-autonomous immunity to both vacuolar and cytosolic replicating bacteria (*152, 153*). For example, loss-of-function screens of all 11 family members in mouse

macrophages clearly demonstrated a protective role for mGbp1, mGbp6, mGbp7 and mGbp10 against infection by *M. bovis BCG* and *L. monoctyogenes* (*154*). Additionally, mGbp2$^{-/-}$ mice are highly susceptible to *Toxoplasma* infection (*155*). The importance of mGbps is also highlighted by bacterial evasion strategies. For example, highly virulent *T. gondii* restrict mGbp recruitment in a rhoptry protein 16 and 18 (ROP16, ROP18)-dependent manner (*156*). These antagonists have been shown to block IRG oligomerization through phosphorylation of residues in the highly conserved GTPase core. Moreover, mGbp5 was shown to have an essential role in NLRP3 inflammasome assembly and subsequent caspase 11-mediated pyroptosis in a stimulus-dependent manner during *Salmonella* and *Listeria* infection (*157, 158*) These studies clearly establish mGbps as effectors of the cell intrinsic response to a wide array of bacterial pathogens.

The role of human GBP (hGBP) proteins is less clear. There are 11 GBPs in mice, relative to only seven in humans. A large bioinformatic survey of human and mouse GBPs recently revealed that the mouse GBP locus has duplicated and undergone multiple copy number duplication events (*149, 150*), resulting in considerable non-orthology relative to the hGBP complement. The lower number of GBPs in the human genome is concomitant with a marked reduction of human IRG genes (two in humans compared to 20 in mice), which is significant given the regulation of mGbps by mIrg proteins. Moreover, mouse-specific GBP paralogs appear to have diverged functionally, such that translation of mouse studies to understanding the function and importance of human GBPs in cell-intrinsic immunity is non-trivial.

*Signatures of positive selection reveal pathogen-driven evolution in primate GBPs*

Pathogen-driven evolution is a hallmark of proteins at the "frontline" of cell-instrinsic immunity, and is an orthogonal approach to identifying genes with longstanding interactions

with pathogens. The enduring nature of host-pathogen arms races is reflected by accelerated rates of replacement (dN, non-synonymous) relative to silent (dS, synonymous) substitutions, or positive selection (dN/dS > 1), in immunity genes. To determine whether GBPs have played an important role against microbial pathogens in primates, we comprehensively screened 13 hominoid, Old World monkey (OWM) and New World monkey (NWM) genomes for evidence consistent with pathogen-driven evolution. Primate GBP orthologs were identified using the human coding sequence as query input to nBLAST, and orthology was verified by synteny and phylogenetic analysis (**Figure 20A**). Orthologs of human GBP1, GBP2, GBP3, GBP6 and GBP7 were identified in all 12 non-human primate genomes. However, we also found dynamic and extensive reorganization of the GBP locus (**Figure 20B**). For example, gibbons represent the most minimal set of GBPs, lacking GBP2, GBP5 and GBP7. This apparent loss coincides with the inversion of GBP3. In OWMs GBP7 has also been lost, whereas there is no evidence for GBP4 and GBP5 in both OWM and NWM genomes. Conservation of the GBP4 and GBP5 genes in hominids and prosimians indicates that these gene loss events were likely independent. In contrast to gibbons, gene loss in these primate families have been supplemented by gene duplication events that have given rise to novel primate GBPs. For example, a second GBP6 gene (denoted GBP6L) is present in all examined OWM and NWM genomes and unambiguously forms a sister taxa to GBP6 (**Figure 20C**). Therefore, GBP6L has been conserved in OWMs and NWMs since its genesis 36-50 million years ago, but has been lost in hominoids. The GBP locus has been more extensively reorganized in NWMs, with tandem duplications of GBP2 and GBP5, representing the most extended GBP complement in primates. In total, there have been eight loss and three gene gain events during the dynamic diversification of GBPs over the course of primate evolution (**Figure 20D**).

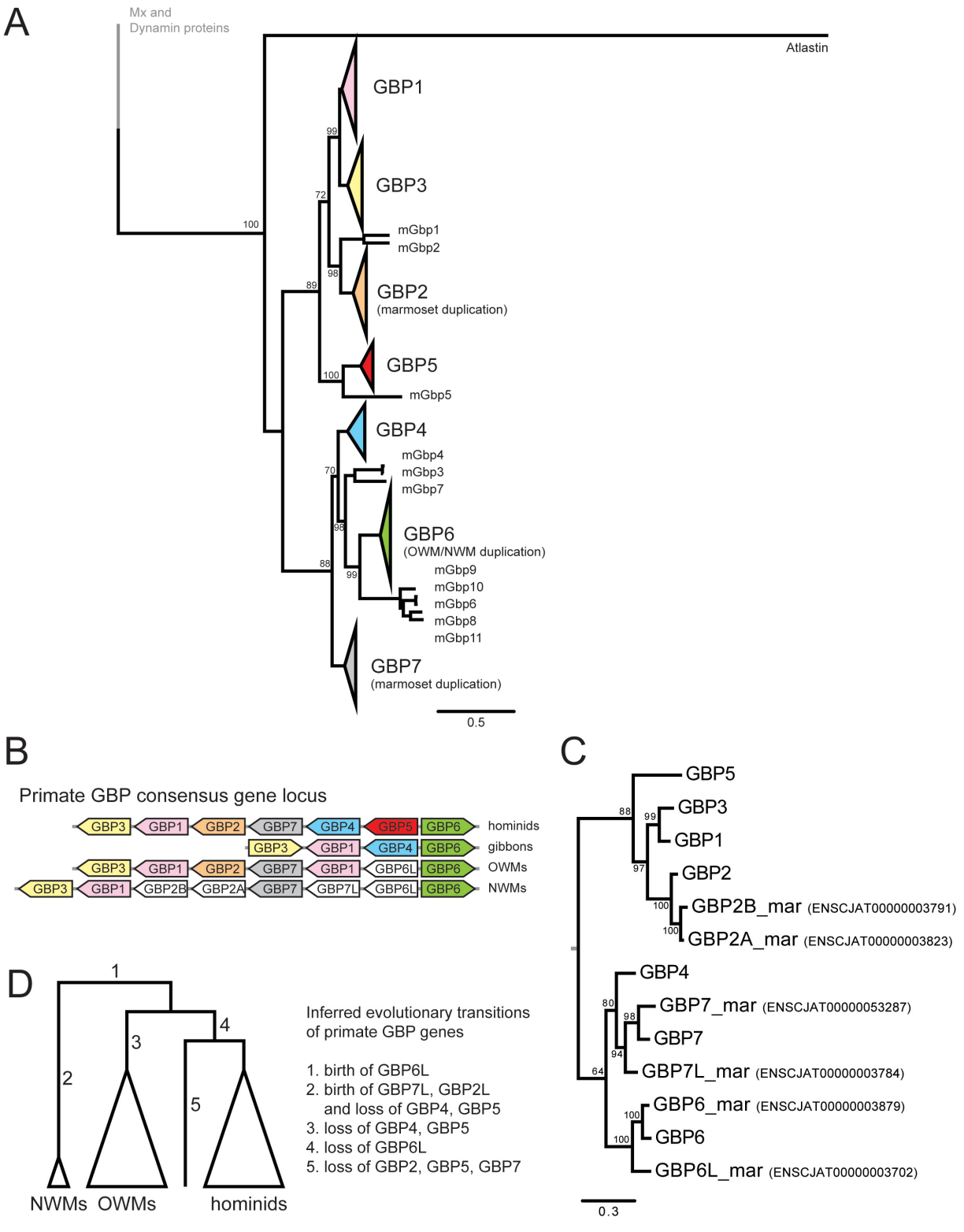

A
Mx and
Dynamin proteins
Atlastin
GBP1
GBP3
mGbp1
mGbp2
GBP2
(marmoset duplication)
GBP5
mGbp5
GBP4
mGbp4
mGbp3
mGbp7
GBP6
(OWM/NWM duplication)
mGbp9
mGbp10
mGbp6
mGbp8
mGbp11
GBP7
(marmoset duplication)
0.5
B
Primate GBP consensus gene locus
GBP3 GBP1 GBP2 GBP7 GBP4 GBP5 GBP6 hominids
GBP3 GBP1 GBP4 GBP6 gibbons
GBP3 GBP1 GBP2 GBP7 GBP1 GBP6L GBP6 OWMs
GBP3 GBP1 GBP2B GBP2A GBP7 GBP7L GBP6L GBP6 NWMs
C
GBP5
GBP3
GBP1
GBP2
GBP2B_mar (ENSCJAT00000003791)
GBP2A_mar (ENSCJAT00000003823)
GBP4
GBP7_mar (ENSCJAT00000053287)
GBP7
GBP7L_mar (ENSCJAT00000003784)
GBP6_mar (ENSCJAT00000003879)
GBP6
GBP6L_mar (ENSCJAT00000003702)
0.3
D
NWMs OWMs hominids
Inferred evolutionary transitions
of primate GBP genes
1. birth of GBP6L
2. birth of GBP7L, GBP2L
and loss of GBP4, GBP5
3. loss of GBP4, GBP5
4. loss of GBP6L
5. loss of GBP2, GBP5, GBP7

**Figure 20. Survey of GBPs in 13 publicly available primate genomes**

**A.** Phylogenetic relationships of all 'canonical' GBPs from 13 publicly available primate genomes. Phylogeny was generated based on protein alignment using PhyML and TreeDyn software (with 1000 bootstrap replicates) (*159*), and rooted on DNM2 and Mx proteins. Primate-specific gene duplications are listed. **B.** Arrangement of GBP locus from representative primates. **C.** Phylogeny of marmoset-specific GBP2, GBP6 and GBP7 gene duplications, based on an alignment of human and marmoset GBP protein sequences. Tree was generated as described for **Figure 20A**. **D.** Evolutionary transitions of primate GBP gene gain and loss.

---

## Table 3. PAML analysis for positive selection in primate DRP

| gene | # orthologs | 2 (ln λ) | *P* value |
|---|---|---|---|
| *House-keeping DRPs* | | | |
| ATL1 | 13 | 0.08 | 0.96 |
| ATL2 | 13 | 1.21 | 0.55 |
| ATL3 | 13 | 3.62 | 0.16 |
| DNM1 | | | n.d. |
| DNM1L | 13 | 0.00 | 1.0 |
| DNM2 | 13 | 0.00 | 1.0 |
| DNM3 | 13 | 1.28 | 0.53 |
| MFN1 | 13 | 2.92 | 0.23 |
| MFN2 | 13 | 0.00 | 1.0 |
| OPA1 | 13 | 2.07 | 0.35 |
| *Interferon-induced DRPs* | | | |
| **Mx1** | 13 | **22.06** | **< 0.001** |
| **Mx2** | 13 | **7.4** | **0.02** |
| GBP1 | 13 | 5.28 | 0.07 |
| **GBP2** | 12 | **6.39** | **0.04** |
| **GBP3** | 12 | **23.35** | **< 0.001** |
| **GBP4** | 8 | **8.11** | **0.02** |
| GBP5 | 7 | 1.89 | 0.39 |
| GBP6 | 13 | 3.71 | 0.16 |
| GBP7 | 11 | 0.00 | 1.0 |

[a] Results of ML tests for positive selection, showing twice the log difference between evolutionary models that allow (M8) or disallow (M7) positive selection.
[b] *P* values describe whether either model was a significantly better fit to the data. Genes for which $P < 0.5$ are in **bold** type.

To determine whether primate GBPs have been subject to pathogen-driven evolution, maximum-likelihood tests were used to assess the rate of non-synonymous relative to synonymous substitutions (dN/dS). Given the minimal dataset for this analysis (10-13 orthologs), we first evaluated the Mx genes, which have been previously reported to evolve under strong positive selection using a more robust set of primate sequences (*102*). Consistent with previous results, both Mx genes showed evidence for positive selection in this dataset (**Table 3**). In

contrast, there was no evidence for positive selection in DRP genes that function outside of cell-intrinsic immunity (**Table 3**), a result in line with expectations based on their highly conserved biological function. Motivated by these findings, we investigated whether primate GBP genes (for which at least 10 orthologs are represented in primates) show evidence of recurrent positive selection. These tests revealed that GBP2, GBP3 and GBP4 have statistically significant signatures of positive selection ($P < 0.05$) (**Table 3**). GBP1 also shows a trend toward significance ($P = 0.07$). We note that the signature of positive selection in GBP3 is particularly striking. Thirteen codons located exclusively in the GBP3 G domain (8/13) and the middle domain of the stalk (5/13, green), as opposed to the GED (0/13, blue) showed evidence for positive selection (PP > 0.90) (**Figure 21A**).

*Evolutionary insights into primate GBP function*

Together, these studies provide evidence for pathogen-driven evolution in primate GBPs. Given the relatively small orthologous gene set used in this study, our analysis likely underestimates the role of positive selection in primate GBP evolution, and motivates deeper inquiries into primate GBP evolution. Importantly, the signature of positive selection in primate GBP1, GBP2 and especially GBP3 prioritize these proteins for functional studies. To date, functional knowledge of hGBP function is almost exclusively via inference from mouse studies. The mouse genome contains two mGbp loci, each of which has undergone extensive diversification. Mice have five copies of hGBP6 (mGbp6, mGbp8, mGbp9, mGbp10 and mGbp11), whereas mGbp3, mGbp4 and mGbp7 are also closely related but distinct from hGBP4 and hGBP6, but are non-orthologous to hGBP7 or hGBP3 (**Figure 20A**). Similarly, mGbp1 and

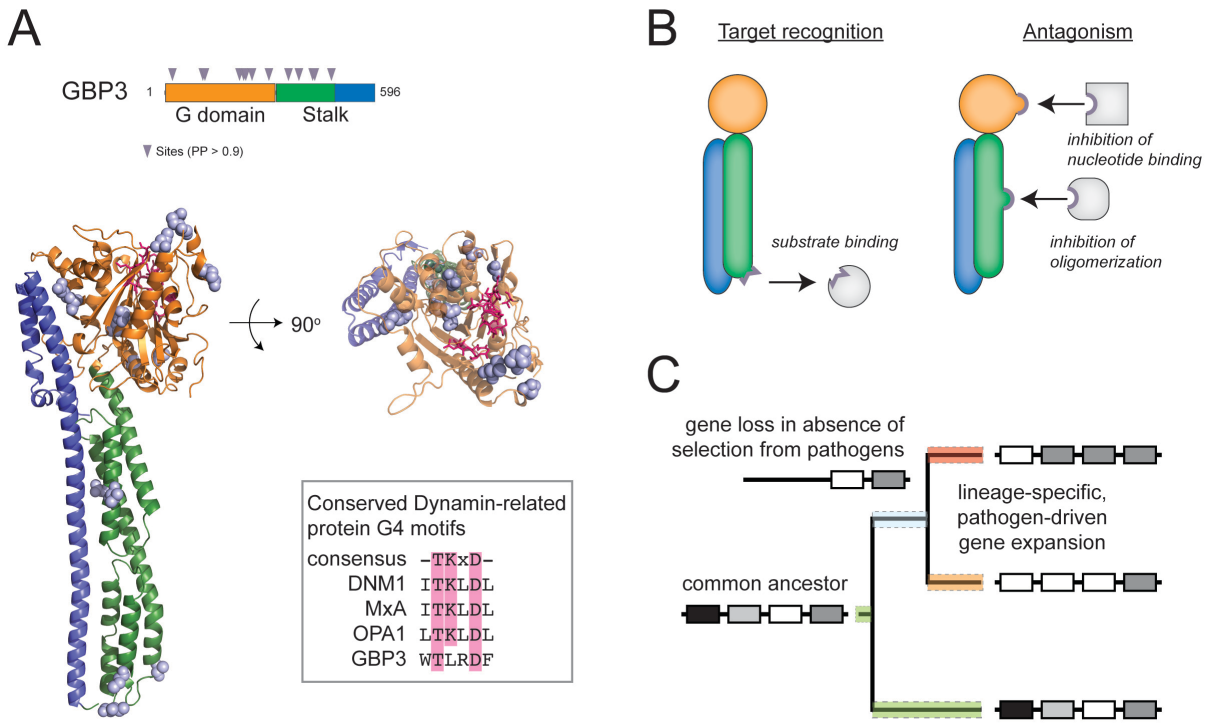



**Figure 21. A model for the pathogen-driven evolution of primate GBP3**

**A.** (upper) A linear schematic of the GBP3 protein with positively selected sites (dN/dS > 0.1, PP > 0.9) marked by triangles. Colors correspond to G domain (orange) and stalk (green). The middle domain (blue) is represented based on conventional nomenclature. (lower) A model of the GBP3 crystal structure built using PHYRE2 (*160*), based largely upon the solved crystal structure of GBP1 (pdb 1DG3 and 1F5N) (*161, 162*). Positively selected sites are shown as surface representations. An alignment of the conserved G4 motif is shown for representative DRPs) (inset). **B.** Model of GBP3-pathogen arms race. Positive selection in the GBP3 unstructured loop between the stalk and middle domain is modeled as a target recognition element, whereas other sites are proposed to reflect evasion from pathogen antagonism of GBP3 essential processes. Color scheme is the same as **Figure 21A**. **C.** Hypothetical model for pathogen-driven host antiviral gene gain and loss events. Host genes are represented as boxes that either expand or retract in a lineage-specific fashion in response to the presence (other colored bars) or absence (blue bar) of pathogens.

mGbp2 are both orthologs of hGBP2 but not hGBP1 (**Figure 20A**). The dynamic and recurrent expansion and contraction of the GBP locus is not restricted to humans and mice. Notably, cows have at least 10 GBP genes while carnivores appear to only encode three. This variability precludes a facile determination of the ancestral state. Hence the "directionality" of GBP evolution in humans and mice (i.e., contraction in humans versus expansion in mice) is difficult to discern in the absence of more sequencing. With these limitations and nomenclature

inconsistencies notwithstanding, true orthologs of hGBP1, hGBP3, hGBP4 and hGBP7 do not exist in the mouse genome. Likewise, functionally characterized mGbps (e.g., mGbp1 and mGbp2; mGbp6, mGbp7 and mGbp10) are copy number variants that may not fully reflect the function of the single gene ortholog in humans. The evolutionary relationship (or lack thereof) between human and mouse GBPs raises important considerations in designing model systems that appropriately reflect human gene function. The recent development of mice lacking an entire mGbp locus provides a platform to conduct "add back" studies with human GBPs to test their sufficiency to confer host resistance to pathogens. Ultimately, the conjunction of evolutionary and functional axes will be required to determine how evolutionary divergence in GBPs influences host susceptibility to microbial pathogens.

Our evolutionary analyses have highlighted GBP3 as a major player in primate immunity. The pattern of positively selected sites in GBP3 indicates that both the G domain and stalk have been strongly shaped by arms races with microbial pathogens. The mechanism by which GBPs engage microbial targets is unclear. GBP associates with bacterial vacuoles through recognition of "self" and "non-self" marks deposited by mIrgm and GKS-containing IRG proteins, respectively (*163*). Although the surfaces involved in GBP-Irgm-IRG interactions have not been fully mapped, such host protein-protein relationships are expected to co-evolve under strong purifying selection. Instead, our data suggest that GBP3 function may also require direct targeting of a pathogen-associated substrate. We note that sites clustered in the disordered loop at the base of the stalk are reminiscent of positive selection in the loop L4 of MxA, which directs MxA target interaction (*102*). Under this model, positive selection reflects GBP3 "chasing" a counter-evolving microbial target (**Figure 21B**).

In contrast, the preponderance of rapidly evolving sites in the G domain and middle stalk are surprising given the highly conserved functions of those domains. For example, two positively selected sites in GBP3 are directly adjacent to the base-coordinating G4 motif (**Figure 21A**). Like all other DRPs - the single exception of MxB notwithstanding - nucleotide binding and homo-oligomerization are essential for GBP anti-pathogen activity. Rapid evolution in regions that coordinate essential aspects of GBP structure and function may forecast an "Achilles heal," which pathogens may exploit to evade restriction. These patterns of positive selection in GBP3 rationalize an evolutionary and molecular model of host-pathogen interaction that provides testable hypotheses for scrutiny in an experimental setting (**Figure 21B**).

The dynamic reorganization of the primate GBP locus raises a number of questions regarding the unique complements of primate and murine GBPs. For example, while the overall copy number of primate GBPs has remained relatively static, there is clear variation in gene content. This is most apparent in gibbons, which have lost nearly half of their GBP repertoire. GBP loss in *Nomascus leucogenys* may be due to severe bottlenecks as the result of human encroachment on habitation and subsequent geographical isolation. Alternatively, the contraction of GBPs in the gibbon genome may speak to functional redundancy encoded by this multigene family. However, this is at odds with the newly established and essential role for human and mouse GBP5, but not other GBPs, in NLRP3 inflammasome assembly (NLRP3 and downstream effectors are conserved in gibbons) (*157, 158*). While the functional consequence of these gene loss events awaits characterization, this highlights the species-specific nature of cell-intrinsic immunity, which may have profound consequences for host susceptibility to microbial infection.

What explains the dramatic gene turnover in the primate and mouse GBP locus? One intriguing possibility is that GBP expansion represents adaptive transitions where each gene has

evolved unique pathogen specificity. Like rapid evolution by positive selection, gene turnover (i.e., copy number expansion and contraction) is another form of genetic innovation that is often observed in genes that function in intrinsic immunity (e.g., Apobecs and TRIMs) (*164-167*). Indeed, neo- or sub-functionalization upon gene duplication is an attractive evolutionary strategy to expand host anti-pathogen repertoires to offset the imbalance between pathogen-specific ISGs and microbial diversity. However, as most ISGs singularly function in host defense, their preservation in host genomes is only guaranteed by ongoing pathogen-driven selective pressure for retention. Therefore, GBP loss may reflect periods of pathogen extinction (**Figure 21C**, blue bar). It has been proposed that gene loss may be accelerated if there is an associated cost to host fitness (e.g., off-target mutation of the host genome via Apobec-mediated cytidine deamination) (*168-170*). Over time the unique emergence or reemergence of pathogens in each lineage would select for the retention of GBP variants with novel pathogen-specificity (**Figure 21C**, red and orange bars). This 'revolving door' pattern of gene evolution in combination with evolutionary signatures of recurrent positive selection reflect host innovations from repeated bouts with counter-adapting pathogens, and highlight the primacy of these genes for host survival. Taken together, the evolutionary studies described herein reflect the ebb and flow of selective pressure from pathogens during primate and rodent evolution.

**References**


1. P. S. Mitchell, M. Emerman, H. S. Malik, An evolutionary perspective on the broad antiviral specificity of MxA. *Curr Opin Microbiol*, (2013).
2. C. A. Janeway, Jr., Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor symposia on quantitative biology* **54 Pt 1**, 1-13 (1989).
3. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
4. L. A. Marraffini, E. J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845 (2008).
5. E. V. Koonin, K. S. Makarova, CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA biology* **10**, 679-686 (2013).
6. H. Li, W. X. Li, S. W. Ding, Induction and suppression of RNA silencing by an animal virus. *Science* **296**, 1319-1321 (2002).
7. D. C. Baulcombe, RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant molecular biology* **32**, 79-88 (1996).
8. M. H. Kumagai, J. Donson, G. della-Cioppa, D. Harvey, K. Hanley, L. K. Grill, Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 1679-1683 (1995).
9. C. A. Janeway, Jr., R. Medzhitov, Innate immune recognition. *Annual review of immunology* **20**, 197-216 (2002).

10. A. Iwasaki, A virological view of innate immune recognition. *Annu Rev Microbiol* **66**, 177-196 (2012).

11. L. Unterholzner, S. E. Keating, M. Baran, K. A. Horan, S. B. Jensen, S. Sharma, C. M. Sirois, T. Jin, E. Latz, T. S. Xiao, K. A. Fitzgerald, S. R. Paludan, A. G. Bowie, IFI16 is an innate immune sensor for intracellular DNA. *Nature immunology* **11**, 997-1004 (2010).

12. L. Sun, J. Wu, F. Du, X. Chen, Z. J. Chen, Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786-791 (2013).

13. S. M. McWhirter, R. Barbalat, K. M. Monroe, M. F. Fontana, M. Hyodo, N. T. Joncker, K. J. Ishii, S. Akira, M. Colonna, Z. J. Chen, K. A. Fitzgerald, Y. Hayakawa, R. E. Vance, A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. *The Journal of experimental medicine* **206**, 1899-1911 (2009).

14. J. W. Schoggins, S. J. Wilson, M. Panis, M. Y. Murphy, C. T. Jones, P. Bieniasz, C. M. Rice, A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481-485 (2011).

15. D. Goubau, S. Deddouche, C. Reis e Sousa, Cytosolic sensing of viruses. *Immunity* **38**, 855-869 (2013).

16. U. Muller, U. Steinhoff, L. F. Reis, S. Hemmi, J. Pavlovic, R. M. Zinkernagel, M. Aguet, Functional role of type I and type II interferons in antiviral defense. *Science* **264**, 1918-1921 (1994).

17. A. Casrouge, S. Y. Zhang, C. Eidenschenk, E. Jouanguy, A. Puel, K. Yang, A. Alcais, C. Picard, N. Mahfoufi, N. Nicolas, L. Lorenzo, S. Plancoulaine, B. Senechal, F. Geissmann, K. Tabeta, K. Hoebe, X. Du, R. L. Miller, B. Heron, C. Mignot, T. B. de Villemeur, P.

Lebon, O. Dulac, F. Rozenberg, B. Beutler, M. Tardieu, L. Abel, J. L. Casanova, Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science* **314**, 308-312 (2006).

18. M. F. Dubois, A. G. Hovanessian, Modified subcellular localization of interferon-induced p68 kinase during encephalomyocarditis virus infection. *Virology* **179**, 591-598 (1990).
19. A. G. Bowie, L. Unterholzner, Viral evasion and subversion of pattern-recognition receptor signalling. *Nat Rev Immunol* **8**, 911-922 (2008).
20. V. D. Menachery, A. J. Eisfeld, A. Schafer, L. Josset, A. C. Sims, S. Proll, S. Fan, C. Li, G. Neumann, S. C. Tilton, J. Chang, L. E. Gralinski, C. Long, R. Green, C. M. Williams, J. Weiss, M. M. Matzke, B. J. Webb-Robertson, A. A. Schepmoes, A. K. Shukla, T. O. Metz, R. D. Smith, K. M. Waters, M. G. Katze, Y. Kawaoka, R. S. Baric, Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses. *mBio* **5**, e01174-01114 (2014).
21. Y. J. Crow, Type I interferonopathies: Mendelian type I interferon up-regulation. *Current opinion in immunology* **32C**, 7-12 (2014).
22. A. J. Sadler, B. R. Williams, Structure and function of the protein kinase R. *Curr Top Microbiol Immunol* **316**, 253-292 (2007).
23. D. Perez-Caballero, T. Zang, A. Ebrahimi, M. W. McNatt, D. A. Gregory, M. C. Johnson, P. D. Bieniasz, Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* **139**, 499-511 (2009).
24. M. G. Grütter, J. Luban, TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. *Curr Opin Virol* **2**, 142-150 (2012).
25. S. P. Goff, Genetic control of retrovirus susceptibility in mammalian cells. *Annu Rev Genet* **38**, 61-85 (2004).

26. C. A. Kozak, A. Chakraborti, Single amino acid changes in the murine leukemia virus capsid protein gene define the target of Fv1 resistance. *Virology* **225**, 300-305 (1996).

27. N. R. Meyerson, S. L. Sawyer, Two-stepping through time: mammals and viruses. *Trends Microbiol* **19**, 286-294 (2011).

28. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* **12**, 687-695 (2012).

29. H. W. Chang, J. C. Watson, B. L. Jacobs, The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4825-4829 (1992).

30. M. V. Davies, O. Elroy-Stein, R. Jagus, B. Moss, R. J. Kaufman, The vaccinia virus K3L gene product potentiates translation by inhibiting double-stranded-RNA-activated protein kinase and phosphorylation of the alpha subunit of eukaryotic initiation factor 2. *J Virol* **66**, 1943-1950 (1992).

31. M. Dube, B. B. Roy, P. Guiot-Guillain, J. Binette, J. Mercier, A. Chiasson, E. A. Cohen, Antagonism of tetherin restriction of HIV-1 release by Vpu involves binding and sequestration of the restriction factor in a perinuclear compartment. *PLoS Pathog* **6**, e1000856 (2010).

32. S. J. Neil, T. Zang, P. D. Bieniasz, Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425-430 (2008).

33. M. Stremlau, C. M. Owens, M. J. Perron, M. Kiessling, P. Autissier, J. Sodroski, The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).

34. M. D. Daugherty, H. S. Malik, Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet* **46**, 677-700 (2012).

35. O. Haller, G. Kochs, Human MxA protein: an interferon-induced dynamin-like GTPase with broad antiviral activity. *J Interferon Cytokine Res* **31**, 79-87 (2011).

36. A. Isaacs, J. Lindenmann, Virus interference. I. The interferon. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society* **147**, 258-267 (1957).

37. J. Lindenmann, Resistance of mice to mouse-adapted influenza A virus. *Virology* **16**, 203-204 (1962).

38. J. Lindenmann, Inheritance of Resistance to Influenza Virus in Mice. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* **116**, 506-509 (1964).

39. O. Haller, H. Arnheiter, I. Gresser, J. Lindenmann, Genetically determined, interferon-dependent resistance to influenza virus in mice. *The Journal of experimental medicine* **149**, 601-612 (1979).

40. P. Staeheli, O. Haller, W. Boll, J. Lindenmann, C. Weissmann, Mx protein: constitutive expression in 3T3 cells transformed with cloned Mx cDNA confers selective resistance to influenza virus. *Cell* **44**, 147-158 (1986).

41. P. Staeheli, O. Haller, Interferon-induced human protein with homology to protein Mx of influenza virus-resistant mice. *Molecular and cellular biology* **5**, 2150-2153 (1985).

42. M. Aebi, J. Fah, N. Hurt, C. E. Samuel, D. Thomis, L. Bazzigher, J. Pavlovic, O. Haller, P. Staeheli, cDNA structures and regulation of two interferon-induced human Mx proteins. *Molecular and cellular biology* **9**, 5062-5072 (1989).

43. M. A. Horisberger, P. Staeheli, O. Haller, Interferon induces a unique protein in mouse cells bearing a gene for resistance to influenza virus. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 1910-1914 (1983).

44. J. Pavlovic, T. Zurcher, O. Haller, P. Staeheli, Resistance to influenza virus and vesicular stomatitis virus conferred by expression of human MxA protein. *J Virol* **64**, 3370-3375 (1990).

45. Z. Liu, Q. Pan, S. Ding, J. Qian, F. Xu, J. Zhou, S. Cen, F. Guo, C. Liang, The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe* **14**, 398-410 (2013).

46. C. Goujon, O. Moncorge, H. Bauby, T. Doyle, C. C. Ward, T. Schaller, S. Hue, W. S. Barclay, R. Schulz, M. H. Malim, Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559-562 (2013).

47. M. Kane, S. S. Yadav, J. Bitzegeio, S. B. Kutluay, T. Zang, S. J. Wilson, J. W. Schoggins, C. M. Rice, M. Yamashita, T. Hatziioannou, P. D. Bieniasz, MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563-566 (2013).

48. H. P. Hefti, M. Frese, H. Landis, C. Di Paolo, A. Aguzzi, O. Haller, J. Pavlovic, Human MxA protein protects mice lacking a functional alpha/beta interferon system against La crosse virus and other lethal viral infections. *J Virol* **73**, 6984-6991 (1999).

49. S. Gao, A. von der Malsburg, A. Dick, K. Faelber, G. F. Schröder, O. Haller, G. Kochs, O. Daumke, Structure of myxovirus resistance protein a reveals intra- and intermolecular domain interactions required for the antiviral function. *Immunity* **35**, 514-525 (2011).

50. .

51. M. L. Rennie, S. A. McKelvie, E. M. Bulloch, R. L. Kingston, Transient dimerization of human MxA promotes GTP hydrolysis, resulting in a mechanical power stroke. *Structure* **22**, 1433-1445 (2014).

52. M. Nakayama, K. Nagata, A. Kato, A. Ishihama, Interferon-inducible mouse Mx1 protein that confers resistance to influenza virus is GTPase. *J Biol Chem* **266**, 21404-21408 (1991).

53. G. J. Praefcke, H. T. McMahon, The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat Rev Mol Cell Biol* **5**, 133-147 (2004).

54. S. Gao, A. von der Malsburg, S. Paeschke, J. Behlke, O. Haller, G. Kochs, O. Daumke, Structural basis of oligomerization in the stalk region of dynamin-like MxA. *Nature* **465**, 502-506 (2010).

55. J. A. Mears, P. Ray, J. E. Hinshaw, A corkscrew model for dynamin constriction. *Structure* **15**, 1190-1202 (2007).

56. A. von der Malsburg, I. Abutbul-Ionita, O. Haller, G. Kochs, D. Danino, Stalk domain of the dynamin-like MxA GTPase protein mediates membrane binding and liposome tubulation via the unstructured L4 loop. *J Biol Chem* **286**, 37858-37865 (2011).

57. K. Faelber, M. Held, S. Gao, Y. Posor, V. Haucke, F. Noé, O. Daumke, Structural insights into dynamin-mediated membrane fission. *Structure* **20**, 1621-1628 (2012).

58. O. Haller, S. Gao, A. von der Malsburg, O. Daumke, G. Kochs, Dynamin-like MxA GTPase: structural insights into oligomerization and implications for antiviral activity. *J Biol Chem* **285**, 28419-28424 (2010).

59. O. Daumke, S. Gao, A. von der Malsburg, O. Haller, G. Kochs, Structure of the MxA stalk elucidates the assembly of ring-like units of an antiviral module. *Small GTPases* **1**, 62-64 (2010).

60. J. Verhelst, E. Parthoens, B. Schepens, W. Fiers, X. Saelens, Interferon-inducible protein Mx1 inhibits influenza virus by interfering with functional viral ribonucleoprotein complex assembly. *J Virol* **86**, 13445-13455 (2012).

61. L. Johannes, R. Kambadur, H. Lee-Hellmich, C. A. Hodgkinson, H. Arnheiter, E. Meier, Antiviral determinants of rat Mx GTPases map to the carboxy-terminal half. *J Virol* **71**, 9792-9795 (1997).

62. P. Zimmermann, B. Mänz, O. Haller, M. Schwemmle, G. Kochs, The viral nucleoprotein determines Mx sensitivity of influenza A viruses. *J Virol* **85**, 8133-8140 (2011).

63. J. Dittmann, S. Stertz, D. Grimm, J. Steel, A. García-Sastre, O. Haller, G. Kochs, Influenza A virus strains differ in sensitivity to the antiviral action of Mx-GTPase. *J Virol* **82**, 3624-3631 (2008).

64. B. Mänz, D. Dornfeld, V. Götz, R. Zell, P. Zimmermann, O. Haller, G. Kochs, M. Schwemmle, Pandemic Influenza A Viruses Escape from Restriction by Human MxA through Adaptive Mutations in the Nucleoprotein. *PLoS Pathog* **9**, e1003279 (2013).

65. D. Riegger, R. Hai, D. Dornfeld, B. Manz, V. Leyva-Grado, M. T. Sanchez-Aparicio, R. A. Albrecht, P. Palese, O. Haller, M. Schwemmle, A. Garcia-Sastre, G. Kochs, M. Schmolke, The nucleoprotein of newly emerged H7N9 influenza A virus harbors a unique motif conferring resistance to antiviral human MxA. *J Virol*, (2014).

66. J. Pavlovic, H. Arzet, H. P. Hefti, M. Frese, D. Rost, B. Ernst, E. Kolb, P. Peter Staeheli, O. Haller, Enhanced Virus Resistance of Transgenic Mice Expressing the Human MxA Protein. 1-5 (1995).

67. G. Kochs, C. Janzen, H. Hohenberg, O. Haller, Antivirally active MxA protein sequesters La Crosse virus nucleocapsid protein into perinuclear complexes. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3153-3158 (2002).

68. G. Kochs, O. Haller, GTP-bound human MxA protein interacts with the nucleocapsids of Thogoto virus (Orthomyxoviridae). *J Biol Chem* **274**, 4370-4376 (1999).

69. G. Kochs, O. Haller, Interferon-induced human MxA GTPase blocks nuclear import of Thogoto virus nucleocapsids. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2082-2086 (1999).

70. J. Pavlovic, O. Haller, P. Staeheli, Human and mouse Mx proteins inhibit different steps of the influenza virus multiplication cycle. *J Virol* **66**, 2564-2569 (1992).

71. P. Staeheli, J. Pavlovic, Inhibition of vesicular stomatitis virus mRNA synthesis by human MxA protein. *J Virol* **65**, 4498-4501 (1991).

72. E. Gordien, O. Rosmorduc, C. Peltekian, F. Garreau, C. Bréchot, D. Kremsdorf, Inhibition of hepatitis B virus replication by the interferon-inducible MxA protein. *Journal of Virology* **75**, 2684-2691 (2001).

73. N. Li, L. Zhang, L. Chen, W. Feng, Y. Xu, F. Chen, X. Liu, Z. Chen, W. Liu, MxA inhibits hepatitis B virus replication by interaction with core protein HBcAg. *Hepatology*, (2012).

74. S. C. Johnston, K. L. Lin, J. H. Connor, G. Ruthel, A. Goff, L. E. Hensley, In vitro inhibition of monkeypox virus production and spread by Interferon-β. *Virol J* **9**, 5 (2012).

75. C. L. Netherton, J. Simpson, O. Haller, T. E. Wileman, H. H. Takamatsu, P. Monaghan, G. Taylor, Inhibition of a large double-stranded DNA virus by MxA protein. *J Virol* **83**, 2310-2320 (2009).

76. H. Landis, A. Simon-Jödicke, A. Klöti, C. Di Paolo, J. J. Schnorr, S. Schneider-Schaulies, H. P. Hefti, J. Pavlovic, Human MxA protein confers resistance to Semliki Forest virus and inhibits the amplification of a Semliki Forest virus-based replicon in the absence of viral structural proteins. *Journal of Virology* **72**, 1516-1522 (1998).

77. D. B. Searls, Pharmacophylogenomics: genes, evolution and drug targets. *Nature reviews. Drug discovery* **2**, 613-623 (2003).

78. S. M. Hedrick, The acquired immune system: a vantage from beneath. *Immunity* **21**, 607-615 (2004).

79. L. Van Valen, A new evolutionary law. *Evolutionary Theory* **1**, 1–30 (1973).

80. C. Kosiol, T. Vinar, R. R. da Fonseca, M. J. Hubisz, C. D. Bustamante, R. Nielsen, A. Siepel, Patterns of positive selection in six Mammalian genomes. *PLoS Genet* **4**, e1000144 (2008).

81. W. Gu, M. Li, Y. Xu, T. Wang, J. H. Ko, T. Zhou, The impact of RNA structure on coding sequence evolution in both bacteria and eukaryotes. *BMC evolutionary biology* **14**, 87 (2014).

82. W. Ran, D. M. Kristensen, E. V. Koonin, Coupling between protein level selection and codon usage optimization in the evolution of bacteria and archaea. *mBio* **5**, e00956-00914 (2014).

83. A. B. Stergachis, E. Haugen, A. Shafer, W. Fu, B. Vernot, A. Reynolds, A. Raubitschek, S. Ziegler, E. M. LeProust, J. M. Akey, J. A. Stamatoyannopoulos, Exonic transcription

factor binding directs codon choice and affects protein evolution. *Science* **342**, 1367-1372 (2013).

84. S. L. Sawyer, L. I. Wu, M. Emerman, H. Malik, Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2832-2837 (2005).

85. E. S. Lim, O. I. Fregoso, C. O. McCoy, F. A. Matsen, H. S. Malik, M. Emerman, The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe* **11**, 194-204 (2012).

86. M. R. Patel, Y. M. Loo, S. M. Horner, M. Gale, H. S. Malik, Convergent evolution of escape from hepaciviral antagonism in primates. *PLoS Biol* **10**, e1001282 (2012).

87. A. A. Compton, V. M. Hirsch, M. Emerman, The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe* **11**, 91-98 (2012).

88. M. D. Daugherty, J. M. Young, J. A. Kerns, H. S. Malik, Rapid evolution of PARP genes suggests a broad role for ADP-ribosylation in host-virus conflicts. *PLoS Genet* **10**, e1004403 (2014).

89. C. J. Spragg, M. Emerman, Antagonism of SAMHD1 is actively maintained in natural infections of simian immunodeficiency virus. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 21136-21141 (2013).

90. O. I. Fregoso, J. Ahn, C. Wang, J. Mehrens, J. Skowronski, M. Emerman, Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the restriction factor SAMHD1. *PLoS Pathog* **9**, e1003496 (2013).

91. J. L. Tenthorey, E. M. Kofoed, M. D. Daugherty, H. S. Malik, R. E. Vance, Molecular basis for specific recognition of bacterial ligands by NAIP/NLRC4 inflammasomes. *Molecular cell* **54**, 17-29 (2014).

92. N. K. Duggal, H. S. Malik, M. Emerman, The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. *J Virol* **85**, 11361-11371 (2011).

93. E. S. Lim, H. S. Malik, M. Emerman, Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *J Virol* **84**, 7124-7134 (2010).

94. N. C. Elde, S. J. Child, A. P. Geballe, H. S. Malik, Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature* **457**, 485-489 (2009).

95. J. A. Kerns, M. Emerman, H. S. Malik, Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet* **4**, e21 (2008).

96. M. OhAinle, J. A. Kerns, H. S. Malik, M. Emerman, Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *J Virol* **80**, 3853-3862 (2006).

97. S. L. Sawyer, L. I. Wu, M. Emerman, H. S. Malik, Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2832-2837 (2005).

98. S. L. Sawyer, M. Emerman, H. S. Malik, Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* **2**, E275 (2004).

99. I. Busnadiego, M. Kane, S. J. Rihn, H. F. Preugschas, J. Hughes, D. Blanco-Melo, V. P. Strouvelle, T. M. Zang, B. J. Willett, C. Boutell, P. D. Bieniasz, S. J. Wilson, Host and viral determinants of Mx2 antiretroviral activity. *J Virol* **88**, 7738-7752 (2014).

100. M. F. Barber, N. C. Elde, Nutritional immunity. Escape from bacterial iron piracy through rapid evolution of transferrin. *Science* **346**, 1362-1366 (2014).

101. A. Demogines, J. Abraham, H. Choe, M. Farzan, S. L. Sawyer, Dual host-virus arms races shape an essential housekeeping protein. *PLoS Biol* **11**, e1001571 (2013).

102. P. S. Mitchell, C. Patzina, M. Emerman, O. Haller, H. S. Malik, G. Kochs, Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe* **12**, 598-604 (2012).

103. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007).

104. O. Haller, M. Frese, D. Rost, P. A. Nuttall, G. Kochs, Tick-borne thogoto virus infection in mice is inhibited by the orthomyxovirus resistance gene product Mx1. *J Virol* **69**, 2596-2601 (1995).

105. E. A. Dietrich, L. Jones-Engel, S. L. Hu, Evolution of the antiretroviral restriction factor TRIMCyp in Old World primates. *PLoS One* **5**, e14019 (2010).

106. P. Perelman, W. E. Johnson, C. Roos, H. N. Seuánez, J. E. Horvath, M. A. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M. P. Schneider, A. Silva, S. J. O'Brien, J. Pecon-Slattery, A molecular phylogeny of living primates. *PLoS Genet* **7**, e1001342 (2011).

107. A. Ponten, C. Sick, M. Weeber, O. Haller, G. Kochs, Dominant-negative mutants of human MxA protein: domains in the carboxy-terminal moiety are important for oligomerization and antiviral activity. *J Virol* **71**, 2591-2599 (1997).

108. Weber, O. Haller, G. Kochs, MxA GTPase blocks reporter gene expression of reconstituted Thogoto virus ribonucleoprotein complexes. *Journal of Virology* **74**, 560-563 (2000).

109. F. Pitossi, A. Blank, A. Schröder, A. Schwarz, P. Hüssi, M. Schwemmle, J. Pavlovic, P. Staeheli, A functional GTP-binding motif is necessary for antiviral activity of Mx proteins. *J Virol* **67**, 6726-6732 (1993).

110. S. Berlin, L. Qu, X. Li, N. Yang, H. Ellegren, Positive diversifying selection in avian Mx genes. *Immunogenetics* **60**, 689-697 (2008).

111. C. Wisskirchen, T. H. Ludersdorfer, D. A. Müller, E. Moritz, J. Pavlovic, Interferon-induced antiviral protein MxA interacts with the cellular RNA helicases UAP56 and URH49. *J Biol Chem* **286**, 34743-34751 (2011).

112. M. Reichelt, S. Stertz, J. Krijnse-Locker, O. Haller, G. Kochs, Missorting of LaCrosse virus nucleocapsid protein by the interferon-induced MxA GTPase involves smooth ER membranes. *Traffic* **5**, 772-784 (2004).

113. A. Le Tortorec, S. Willey, S. J. Neil, Antiviral inhibition of enveloped virus release by tetherin/BST-2: action and counteraction. *Viruses* **3**, 520-540 (2011).

114. J. W. Schoggins, D. A. MacDuff, N. Imanaka, M. D. Gainey, B. Shrestha, J. L. Eitson, K. B. Mar, R. B. Richardson, A. V. Ratushny, V. Litvak, R. Dabelic, B. Manicassamy, J. D. Aitchison, A. Aderem, R. M. Elliott, A. Garcia-Sastre, V. Racaniello, E. J. Snijder, W. M. Yokoyama, M. S. Diamond, H. W. Virgin, C. M. Rice, Pan-viral specificity of IFN-

induced genes reveals new roles for cGAS in innate immunity. *Nature* **505**, 691-695 (2014).

115. N. Biris, Y. Yang, A. B. Taylor, A. Tomashevski, M. Guo, P. J. Hart, F. Diaz-Griffero, D. N. Ivanov, Structure of the rhesus monkey TRIM5alpha PRYSPRY domain, the HIV capsid recognition module. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13278-13283 (2012).

116. A. Kirmaier, F. Wu, R. M. Newman, L. R. Hall, J. S. Morgan, S. O'Connor, P. A. Marx, M. Meythaler, S. Goldstein, A. Buckler-White, A. Kaur, V. M. Hirsch, W. E. Johnson, TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. *PLoS Biol* **8**, (2010).

117. C. Patzina, O. Haller, G. Kochs, Structural requirements for the antiviral activity of the human MxA protein against Thogoto and influenza A virus. *J Biol Chem* **289**, 6020-6027 (2014).

118. M. Schwemmle, K. C. Weining, M. F. Richter, B. Schumacher, P. Staeheli, Vesicular stomatitis virus transcription inhibited by purified MxA protein. *Virology* **206**, 545-554 (1995).

119. B. S. Heinrich, D. K. Cureton, A. A. Rahmeh, S. P. Whelan, Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. *PLoS Pathog* **6**, e1000958 (2010).

120. S. Schneider-Schaulies, J. Schneider-Schaulies, A. Schuster, M. Bayer, J. Pavlovic, V. ter Meulen, Cell type-specific MxA-mediated inhibition of measles virus transcription in human brain cells. *J Virol* **68**, 6910-6917 (1994).

121. J. J. Schnorr, S. Schneider-Schaulies, A. Simon-Jodicke, J. Pavlovic, M. A. Horisberger, V. ter Meulen, MxA-dependent inhibition of measles virus glycoprotein synthesis in a stably transfected human monocytic cell line. *J Virol* **67**, 4760-4768 (1993).

122. M. Vignuzzi, J. K. Stone, R. Andino, Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. *Virus research* **107**, 173-181 (2005).

123. A. S. Lauring, R. Andino, Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* **6**, e1001005 (2010).

124. E. Domingo, J. Sheldon, C. Perales, Viral quasispecies evolution. *Microbiology and molecular biology reviews : MMBR* **76**, 159-216 (2012).

125. M. E. Caines, K. Bichel, A. J. Price, W. A. McEwan, G. J. Towers, B. J. Willett, S. M. Freund, L. C. James, Diverse HIV viruses are targeted by a conformationally dynamic antiviral. *Nat Struct Mol Biol* **19**, 411-416 (2012).

126. B. Mänz, D. Dornfeld, V. Götz, R. Zell, P. Zimmermann, O. Haller, G. Kochs, M. Schwemmle, Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. *PLoS Pathogens*, (2013).

127. N. C. Elde, H. S. Malik, The evolutionary conundrum of pathogen mimicry. *Nat Rev Microbiol* **7**, 787-797 (2009).

128. E. J. Tanner, H. M. Liu, M. S. Oberste, M. Pallansch, M. S. Collett, K. Kirkegaard, Dominant drug targets suppress the emergence of antiviral resistance. *eLife* **3**, (2014).

129. M. T. Ferris, D. L. Aylor, D. Bottomly, A. C. Whitmore, L. D. Aicher, T. A. Bell, B. Bradel-Tretheway, J. T. Bryan, R. J. Buus, L. E. Gralinski, B. L. Haagmans, L. McMillan, D. R. Miller, E. Rosenzweig, W. Valdar, J. Wang, G. A. Churchill, D. W. Threadgill, S.

K. McWeeney, M. G. Katze, F. Pardo-Manuel de Villena, R. S. Baric, M. T. Heise, Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. *PLoS Pathog* **9**, e1003196 (2013).

130. J. L. Fribourgh, H. C. Nguyen, K. A. Matreyek, F. J. Alvarez, B. J. Summers, T. G. Dewdney, C. Aiken, P. Zhang, A. Engelman, Y. Xiong, Structural Insight into HIV-1 Restriction by MxB. *Cell Host Microbe* **16**, 627-638 (2014).

131. B. Xu, J. Kong, X. Wang, W. Wei, W. Xie, X. F. Yu, Structural insight into the assembly of human anti-HIV dynamin-like protein MxB/Mx2. *Biochemical and biophysical research communications* **456**, 197-201 (2015).

132. L. W. Enquist, V. Editors of the Journal of, Virology in the 21st century. *J Virol* **83**, 5296-5308 (2009).

133. S. C. Johnston, K. L. Lin, J. H. Connor, G. Ruthel, A. Goff, L. E. Hensley, In vitro inhibition of monkeypox virus production and spread by Interferon-beta. *Virol J* **9**, 5 (2012).

134. N. Vasilakis, J. Cardosa, K. A. Hanley, E. C. Holmes, S. C. Weaver, Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nat Rev Microbiol* **9**, 532-541 (2011).

135. J. E. Bryant, E. C. Holmes, A. D. Barrett, Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog* **3**, e75 (2007).

136. M. Emerman, H. Malik, Paleovirology--modern consequences of ancient viruses. *PLoS Biology* **8**, e1000301 (2010).

137. A. Hoenen, W. Liu, G. Kochs, A. A. Khromykh, J. M. Mackenzie, West Nile virus-induced cytoplasmic membrane structures provide partial protection against the interferon-induced antiviral MxA protein. *J Gen Virol* **88**, 3013-3017 (2007).

138. A. Hoenen, L. Gillespie, G. Morgan, P. van der Heide, A. Khromykh, J. Mackenzie, The West Nile virus assembly process evades the conserved antiviral mechanism of the interferon-induced MxA protein. *Virology* **448**, 104-116 (2014).

139. K. Liu, X. Liao, B. Zhou, H. Yao, S. Fan, P. Chen, D. Miao, Porcine alpha interferon inhibit Japanese encephalitis virus replication by different ISGs in vitro. *Research in veterinary science* **95**, 950-956 (2013).

140. S. Stertz, M. Reichelt, J. Krijnse-Locker, J. Mackenzie, J. C. Simpson, O. Haller, G. Kochs, Interferon-induced, antiviral human MxA protein localizes to a distinct subcompartment of the smooth endoplasmic reticulum. *J Interferon Cytokine Res* **26**, 650-660 (2006).

141. A. K. Haldar, H. A. Saka, A. S. Piro, J. D. Dunn, S. C. Henry, G. A. Taylor, E. M. Frickel, R. H. Valdivia, J. Coers, IRG and GBP host resistance factors target aberrant, "non-self" vacuoles characterized by the missing of "self" IRGM proteins. *PLoS Pathog* **9**, e1003414 (2013).

142. T. Zürcher, J. Pavlovic, P. Staeheli, Nuclear localization of mouse Mx1 protein is necessary for inhibition of influenza virus. *J Virol* **66**, 5059-5066 (1992).

143. S. L. Sawyer, N. C. Elde, A cross-species view on viruses. *Curr Opin Virol* **2**, 561-568 (2012).

144. M. C. King, G. Raposo, M. A. Lemmon, Inhibition of nuclear import and cell-cycle progression by mutated forms of the dynamin-like GTPase MxB. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8957-8962 (2004).

145. M. J. de Veer, M. Holko, M. Frevel, E. Walker, S. Der, J. M. Paranjape, R. H. Silverman, B. R. Williams, Functional classification of interferon-stimulated genes identified using microarrays. *Journal of leukocyte biology* **69**, 912-920 (2001).

146. Z. Yang. (2014).

147. Z. Yang, R. Nielsen, Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* **19**, 908-917 (2002).

148. C. Goujon, O. Moncorge, H. Bauby, T. Doyle, W. S. Barclay, M. H. Malim, Transfer of the amino-terminal nuclear envelope targeting domain of human MX2 converts MX1 into an HIV-1 resistance factor. *J Virol* **88**, 9017-9026 (2014).

149. C. Bekpen, J. P. Hunn, C. Rohde, I. Parvanova, L. Guethlein, D. M. Dunn, E. Glowalla, M. Leptin, J. C. Howard, The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. *Genome biology* **6**, R92 (2005).

150. M. A. Olszewski, J. Gray, D. J. Vestal, In silico genomic analysis of the human and murine guanylate-binding protein (GBP) gene clusters. *J Interferon Cytokine Res* **26**, 328-352 (2006).

151. S. Martens, J. Howard, The interferon-inducible GTPases. *Annual review of cell and developmental biology* **22**, 559-589 (2006).

152. B. H. Kim, A. R. Shenoy, P. Kumar, C. J. Bradfield, J. D. MacMicking, IFN-inducible GTPases in host cell defense. *Cell Host Microbe* **12**, 432-444 (2012).

153. D. Degrandi, C. Konermann, C. Beuter-Gunia, A. Kresse, J. Wurthner, S. Kurig, S. Beer, K. Pfeffer, Extensive characterization of IFN-induced GTPases mGBP1 to mGBP10 involved in host defense. *Journal of immunology* **179**, 7729-7740 (2007).

154. B. H. Kim, A. R. Shenoy, P. Kumar, R. Das, S. Tiwari, J. D. MacMicking, A family of IFN-gamma-inducible 65-kD GTPases protects against bacterial infection. *Science* **332**, 717-721 (2011).

155. E. Kravets, D. Degrandi, S. Weidtkamp-Peters, B. Ries, C. Konermann, S. Felekyan, J. M. Dargazanli, G. J. Praefcke, C. A. Seidel, L. Schmitt, S. H. Smits, K. Pfeffer, The GTPase activity of murine guanylate-binding protein 2 (mGBP2) controls the intracellular localization and recruitment to the parasitophorous vacuole of Toxoplasma gondii. *J Biol Chem* **287**, 27452-27466 (2012).

156. T. Steinfeldt, S. Konen-Waisman, L. Tong, N. Pawlowski, T. Lamkemeyer, L. D. Sibley, J. P. Hunn, J. C. Howard, Phosphorylation of mouse immunity-related GTPase (IRG) resistance proteins is an evasion strategy for virulent Toxoplasma gondii. *PLoS Biol* **8**, e1000576 (2010).

157. D. M. Pilla, J. A. Hagar, A. K. Haldar, A. K. Mason, D. Degrandi, K. Pfeffer, R. K. Ernst, M. Yamamoto, E. A. Miao, J. Coers, Guanylate binding proteins promote caspase-11-dependent pyroptosis in response to cytoplasmic LPS. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 6046-6051 (2014).

158. A. R. Shenoy, D. A. Wellington, P. Kumar, H. Kassa, C. J. Booth, P. Cresswell, J. D. MacMicking, GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science* **336**, 481-485 (2012).

159. A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie, O. Gascuel, Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research* **36**, W465-469 (2008).

160. L. A. Kelley, M. J. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols* **4**, 363-371 (2009).

161. B. Prakash, G. J. Praefcke, L. Renault, A. Wittinghofer, C. Herrmann, Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* **403**, 567-571 (2000).

162. B. Prakash, L. Renault, G. J. Praefcke, C. Herrmann, A. Wittinghofer, Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *The EMBO journal* **19**, 4555-4564 (2000).

163. J. Coers, Self and non-self discrimination of intracellular membranes by the innate immune system. *PLoS Pathog* **9**, e1003538 (2013).

164. K. Han, D. I. Lou, S. L. Sawyer, Identification of a genomic reservoir for new TRIM genes in primate genomes. *PLoS Genet* **7**, e1002388 (2011).

165. R. Malfavon-Borja, S. L. Sawyer, L. I. Wu, M. Emerman, H. S. Malik, An evolutionary screen highlights canonical and noncanonical candidate antiviral genes within the primate TRIM gene family. *Genome biology and evolution* **5**, 2141-2154 (2013).

166. S. G. Conticello, C. J. Thomas, S. K. Petersen-Mahrt, M. S. Neuberger, Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* **22**, 367-377 (2005).

167. C. Munk, A. Willemsen, I. G. Bravo, An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC evolutionary biology* **12**, 71 (2012).

168. S. Nik-Zainal, D. C. Wedge, L. B. Alexandrov, M. Petljak, A. P. Butler, N. Bolli, H. R. Davies, S. Knappskog, S. Martin, E. Papaemmanuil, M. Ramakrishna, A. Shlien, I. Simonic, Y. Xue, C. Tyler-Smith, P. J. Campbell, M. R. Stratton, Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nature genetics* **46**, 487-491 (2014).

169. S. A. Roberts, M. S. Lawrence, L. J. Klimczak, S. A. Grimm, D. Fargo, P. Stojanov, A. Kiezun, G. V. Kryukov, S. L. Carter, G. Saksena, S. Harris, R. R. Shah, M. A. Resnick, G. Getz, D. A. Gordenin, An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature genetics* **45**, 970-976 (2013).

170. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, I. Australian Pancreatic Cancer Genome, I. B. C. Consortium, I. M.-S. Consortium, I.

PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, M. R. Stratton, Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).