The dynamic proteome of Influenza A virus infection identifies species- and strain-specific control of M segment RNA splicing

vorgelegt von

Dipl.-Ing.

Boris Bogdanow

an der Fakultät III – Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing.-

genehmigte Dissertation

Promotionsausschuss

Vorsitzender: Prof. Dr. Peter Neubauer, TU Berlin

Gutachter: Prof. Dr. Jens Kurreck, TU Berlin

Gutachter: Prof. Dr. Matthias Selbach, MDC Berlin

Tag der wissenschaftlichen Aussprache: 15. Januar 2020

Berlin 2020

~ ~

Vorsitzender:
Prof. Dr. rer. nat. Peter Neubauer
Technische Universität Berlin
Fakultät III, Institut für Biotechnologie, Fachgebiet Bioverfahrenstechnik
Ackerstraße 76, 13355 Berlin

Gutachter 1:
Prof. Dr. rer. nat. Jens Kurreck
Technische Universität Berlin
Fakultät III, Institut für Biotechnologie, Fachgebiet Angewandte Biochemie
Gustav-Meyer-Allee 25, 13355 Berlin

Gutachter 2:
Prof. Dr. rer. nat. Matthias Selbach
Max-Delbrück-Centrum für Molekulare Medizin
Proteome Dynamics, Gebäude 31.1, Raum 3014
Robert-Rössle-Straße 10, 13125 Berlin

for my family

SUMMARY

All viruses hijack host cells to replicate. To this end, they have to interact with cellular factors at all stages of the replication cycle. Since these host cell factors differ markedly between species, viruses typically need to adapt in order to cross a species barrier. Advances in high-throughput sequencing have provided insights into the extraordinary diversity of viruses and genomic determinants of host adaptation. However, the mechanism how these adaptive mutations enable replication in a given host is much less clear.

Proteins are key players during virus infections. The study of many proteins (i.e. the proteome) in parallel was enabled by the advent of mass-spectrometry based proteomics. Further, quantitative proteomics provides the means to assess differences in proteome composition comparing different treatments or cell states.

Here, quantitative proteomics in combination with bioorthogonal labeling was applied to study host-pathogen interaction at the level of protein synthesis. This approach was used to monitor protein synthesis in response to infection with human or avian-adapted Influenza A viruses in human cells. While host proteins were synthesized surprisingly similar between both avian- and human IAV infected cells, the synthesis of some viral proteins, such as the Neuraminidase, Hemagglutinin and Matrix protein M1 was remarkably different. M1 was inefficiently produced by the avian-adapted virus, which was linked to differences in M1 mRNA levels. Experiments with reporter vectors and recombinant viruses show that strain-specific splicing of the M segment pre-mRNA explains partially the inefficient production of M1. Importantly, the splicing differences are

controlled by a *cis*-regulatory element encompassing the 3’ splice site region of the M segment. This element exhibits characteristic and evolutionary conserved differences in RNA secondary structure between avian- and human-adapted IAV strains. Furthermore, exchanging eight nucleotide at the 3’ splice site region impaired multiplication of a human-adapted IAV strain. Taken together, this thesis identifies strain- and species-specific differences in control of viral protein synthesis. The control of M segment RNA splicing was linked to host adaptation. Understanding the host and viral determinants of species-specific M segment RNA splicing may increase our abilities for risk assessment of potential pandemic events of IAV.

ZUSAMMENFASSUNG

Alle Viren befallen Wirtszellen, um sich zu replizieren. Zu diesem Zweck müssen sie in allen Phasen des Replikationszyklus mit zellulären Faktoren interagieren. Da sich diese Faktoren der Wirtszelle von speziesabhängig stark unterscheiden können, müssen sich die Viren typischerweise anpassen, um eine Artenbarriere zu überschreiten. Fortschritte in der Hochdurchsatz-Sequenzierung haben Einblicke in die außergewöhnliche Vielfalt von Viren und deren genomischen Determinanten der Wirtsadaption ermöglicht. Der Mechanismus, wie diese adaptiven Mutationen die Replikation in einem bestimmten Wirt ermöglichen, ist jedoch meist weniger klar.

Proteine sind Schlüsselfaktoren bei Virusinfektionen. Die Untersuchung vieler Proteine (d.h. des gesamten Proteoms) wurde durch das Aufkommen der massenspektrometrisch basierten Proteomik ermöglicht. Darüber hinaus bietet die quantitative Proteomik die Möglichkeit, Unterschiede in der Proteomzusammensetzung im Vergleich verschiedener Behandlungen oder Zellzustände zu beurteilen.

Hier wurde die quantitative Proteomik in Kombination mit bioorthogonaler Markierung eingesetzt, um die Wirt-Pathogen-Interaktion auf der Ebene der Proteinsynthese zu untersuchen. Dieser Ansatz wurde verwendet, um die Proteinsynthese als Reaktion auf eine Infektion mit human- oder vogeladaptierten Influenza A Viren in menschlichen Zellen zu untersuchen. Während die Wirtsproteine zwischen Vogel- und humanen IAV-infizierten Zellen ähnlich synthetisiert wurden, war die Synthese einiger viraler Proteine wie Neuraminidase, Hämagglutinin und Matrixprotein M1 bemerkenswert unterschiedlich. M1 wurde ineffizient vom vogel-adaptierten Stamm produziert. Dies wurde mit

Unterschieden in den M1 mRNA Mengen in Verbindung gebracht. Experimente mit Reportervektoren und rekombinanten Viren zeigen, dass das stammspezifische Spleißen der M-Segment pre-mRNA teilweise die ineffiziente Produktion von M1 erklärt. Die Unterschiede im Spleißen werden durch ein *cis*-regulierendes Element gesteuert, das den 3' Spleißstellenbereich des M-Segments umfasst. Dieses Element weist charakteristische und evolutionär konservierte Unterschiede in der RNA-Sekundärstruktur zwischen vogel- und humanadaptierten IAV-Stämmen auf. Darüber hinaus beeinträchtigte der Austausch von acht Nukleotiden im Bereich der 3'-Spleißstelle die Vermehrung eines humanadaptierten IAV-Stamms. Zusammenfassend zeichnet diese Arbeit ein globales Bild der Proteinneusynthese und Transkription nach Infektion mit Influenza A Viren. Es wurden Stamm- und Spezies-spezifische Unterschiede in der Kontrolle der viralen Proteinsynthese identifiziert. Darüberhinaus geht die Kontrolle des M-Segments RNA-Spleißens mit der Wirtsanpassung von Influenza A Viren einher. Das Verständnis der wirts- und virusabhängigen Determinanten des M Segment RNA Spleißens könnte die Risikoabschätzung prä-pandemischer Influenza A Stämme erleichtern.

STATEMENT OF CONTRIBUTION

The work presented here is built on a collaborative effort, involving many people with different expertise. For clarity the contributions to individual results are listed below. In addition these contributions are mentioned in the corresponding figure legends.

The idea to use AHA-labeling to assess proteome dynamics was contributed by **Katrin Eichelbaum** and **Matthias Selbach** (MDC Berlin). The model system that was used (A549 cells, Pan and Mal viruses) was contributed by **Thorsten Wolff** (RKI Berlin).

Katrin Eichelbaum established the AHA-labeling approach and I collected pAHA and AP-MS mass-spec raw-files with her.

Anne Sadewasser, **Gudrun Heins**, **Katharina Paki** and **Thorsten Wolff** (RKI Berlin) performed and supervised the infections for the pAHA-SILAC exeriment. Further they contributed the multicycle replication curve of Pan, Mal, Pan-AV and Pan + Mal M viruses (Figure 19e).

Irmtraud Meyer (MDC Berlin) contributed the bioinformatics prediction of RNA secondary structure (Figure 18a and Supplementary Figure 6).

Xi Wang, **Jingyi Hou** and **Wei Chen** (MDC Berlin) performed the RNAseq experiments. These data were used for Figures 12b,c, 15a, 16a,b,d and Supplementary Figure 3b.

Barbara Vetter, **Lüder Wiebusch** (Charité Berlin) and **Martha Hergeselle** (MDC Berlin) contributed individual plasmids that were used for transfections in the experiments for Figures 17c and 19d,e.

As part of his studies in Biotechnology **Immanuel Husic** (MDC Berlin) performed qRT-PCR measurements depicted in Figures 16e, 17d and 19d.

SELBSTÄNDIGKEITSERKLÄRUNG

Hiermit versichere ich, Boris Bogdanow, geboren am 26.06.1986, dass ich die vorliegende Dissertation mit dem Titel „The dynamic proteome of Influenza A virus infection identifies species- and strain-specific control of M segment RNA splicing“ selbständig angefertigt und keine weiteren Hilfsmittel und Quellen benutzt habe als die, die im Literaturverzeichnis angegeben sind. Die digitale Version der Arbeit stimmt mit der gedruckten Version überein.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht.

Die Arbeit wurde in gleicher oder ähnlicher Form noch nicht als Prüfungsleistung eingereicht.

Berlin, den …… ……

Boris Bogdanow

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY FIGURES

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Virology

1.1.1. The association of viruses with disease

Viruses are the causative agents for a wide variety of human diseases. The spectrum of clinical manifestations due to virus infections is very broad and ranges from the relatively mild symptoms of the common cold (i.e. rhinoviruses, (Jacobs et al., 2013)), persistence with occasional reactivations (i.e. Herpes Simplex virus, (Roizmann et al., 1992)), life-threatening hemorrhagic fevers (i.e. Marburg virus (Brauburger et al., 2012)) or causing of tumors (i.e. Epstein-Barr virus (Adler et al., 2017)).

The association of diseases with viruses was proposed by numerous researchers in the late 19th century, including Adolf Mayer, Dimitri Ivanovsky and Martinius Beijerinik who were working on Tobacco mosaic virus (Lustig and Levine, 1992). These researchers established that the pathogenic agent of tobacco mosaic disease is so small that it could not be visualized by light microscopy, that it is not retained when the liquid is filtered and that it is able to cause disease by multiplying in living cells (Lustig and Levine, 1992). In 1898, Loeffler and Frosch isolated foot and mouth disease virus of cattle, as the first animal virus, and Walter Reed and colleagues isolated with yellow fever virus the first human virus (Lustig and Levine, 1992).

However, it was difficult to directly and causatively link a virus to an infectious disease. This was a particularly vexing problem in the case of Influenza. In 1892, Richard Pfeiffer reported the isolation of a bacterium that was called “Bacillus influenza” (now: *Haemophilus influenza*) from Influenza patients. The bacteria were present in high

number, could be cultured and transferred to monkeys and rabbits. This led the researcher to mistakenly consider the identified bacterium as the cause for Influenza disease (Pfeiffer, 1892). In 1931 Richard Shope filtered samples from swine that caused a contagious mild form of an influenza-like disease that he called “filtrate disease” (Shope, 1931). He correctly concluded from his studies that a virus causes the disease, which then facilitates secondary infection with *H. influenza.* Briefly after, the human Influenza A virus was isolated, when Smith and colleagues filtered throat washings of patients of the 1933 influenza epidemic through a membrane that retains bacteria. The filtrates were applied to ferrets, which subsequently showed symptoms of Influenza disease (Smith et al., 1995).

By 2012, 219 virus species were known to infect humans and it is estimated that many more will arise (Woolhouse et al., 2012). These viruses are responsible for a significant share of the public health burden. In addition, viruses are increasingly implicated as risk factors in the emergence and pathogenic mechanisms of non-communicable diseases (diseases that are traditionally not thought to be transmittable between humans, such as Alzheimer’s disease or Multiple Sclerosis) (Ogoina and Onyemelukwe, 2009).

1.1.2. Productive and non-productive infections

Viruses can be defined as infective agents that “consist of a nucleic acid molecule in a protein coat and are able to multiply only within the living cells of a host.” (Oxford Dictionaries, 2019). This means that the viral life cycle is composed of two stages, a virion stage and a productive cellular stage. During the former the viral genetic material is packed in proteinaceous shell that confers specificity to the virus. It serves to transport copies of the viral genome from one cell to another. During the cellular stage a host cell is entered and virus gene expression is started, which leads eventually to cell lysis and release of viral progeny (that is, a productive infection outcome). However, the outcome of an infection may not always be productive. Some cells may not be permissive for a virus, which may lead to an abortive infection (Israel, 1980) (i.e. very little or no novel progeny are produced). Also, some viruses (e.g. herpesviruses) have the ability to initiate a latent infection, which is defined as the maintenance of the viral genome in the cell in the absence of productive replication. Also, cells can be infected and produce a low number of viral progeny over a prolonged time period (*persistent* infection).

At the cellular level, the outcome of infection is dependent on certain cellular characteristics. Firstly, the specific cell type may not allow that a virus is productively produced. For example the Hepatitis C virus is practically only able to replicate *in vitro* autonomously in the Huh-7 cell line (Blight et al., 2002). Also, the cell cycle stage can influence whether a productive cycle is initiated. This is for example the case with human cytomegalovirus (HCMV), a virus that induces the production of lytic gene products not during DNA synthesis (S) or gap 2 (G2) phase (Bogdanow et al., 2013).

A non-productive infection may also occur, when a virus is not adapted to a specific host species (e.g. avian Influenza a virus strains in human cells) (Inglis and Brown, 1984; Israel, 1980; Sadewasser et al., 2017).

1.1.3. Virus-host coevolution and host-switching

A fascinating aspect of virology is the diversity of its subject. Viruses differ in their use of genetic material (DNA or RNA), in organization of the genome (linear, segmented, circular), the size of the genome, the size and structure of the virion, replication mechanism and interaction with the cellular host.

Virology traditionally focuses on viruses causing disease in humans, economically relevant animals or plants. With the advent of metagenomics/transcriptomics and high-throughput sequencing numerous hosts could be analyzed for the totality of viruses (that is, the *virome*) (Angly et al., 2006; Li et al., 2015; Shi et al., 2016; Wigington et al., 2016). These data are strikingly changing the way how we think about virus diversity. For example, the currently ICTV (International Committee on Taxonomy of Viruses)-classified 5,560 virus species (International Committee on Taxonomy of Viruses, 2019) are likely only a tiny fraction (much less than 0.1 %) of the entire virosphere (that is, all viruses) (Geoghegan and Holmes, 2017).

The diversity of viruses is intimately connected to the long-term interaction with the cellular host. In principle, viruses and hosts can co-adapt over long evolutionary time scales (co-evolution). This is considered to be the case when phylogenetic trees of viruses mirror that of their hosts. In contrast, host-switching can also occur. Here, the phylogenetic trees of host and virus are in disarray. A recent metatranscriptomic study found for RNA viruses that they are overall co-evolving with their hosts (Shi et al., 2018). However, host-switching events are also frequently observed, especially in the case of Influenza (Shi et al., 2018). The extent to which viruses switch their host or co-evolve with them depends on the virus family. A phylogenetic study of 19 virus

families and their eukaryotic hosts found that double-stranded DNA viruses tend to co-evolve with their hosts to a stronger degree than RNA viruses, which displayed more frequent host-switching (Geoghegan et al., 2017). For example, Herpesviruses are thought to co-evolve with their hosts over millions of years, which makes them highly species-specific. However, this view has only recently been challenged by data gathered through metatranscriptomics/ genomics about entire viromes (Azab et al., 2018).

Importantly, the host-switching of a virus is central to the emergence of a disease. While most of the naturally occurring viruses pose no threat to humans, a tiny fraction of them is known to switch their hosts and cause disease in humans (that is *zoonoses*). Examples are arthropod-borne viruses that can be transmitted to humans via mosquitoes, such as dengue virus (Halstead, 2008), viruses that have their natural reservoir in bats, such as Ebola virus (Malvy et al., 2019) or influenza viruses, which have their natural reservoir in wild waterfowl (Long et al., 2019). A recent estimation states that between ~600,000 and ~800,000 RNA viruses of mammals and birds could infect humans (Carroll et al., 2018).

Fortunately, there are several barriers that hinder most of these animal viruses from becoming *endemic* in humans (that is, the ability of a virus for sustained and efficient replication in and transmission among humans). To attain an endemic state, a virus from an animal reservoir must overcome a series of barriers (Figure 1) (Plowright et al., 2017). This includes first the actual exposure to the pathogen, by e.g. close contact to domestic animals that harbor the pathogen. Second, human cells need to be entered (*susceptible*) and be *permissive* for the virus (i.e. the virus can replicate and

produce new progeny in a given cell type). Lastly, viruses need to attain sustained human-to human transmission to become *endemic*. Collectively, these barriers are a function of a complex interplay between epidemiological, ecological and molecular factors. They filter the huge reservoir of animal viruses so that only very few viruses become relevant for humans in a disease context.

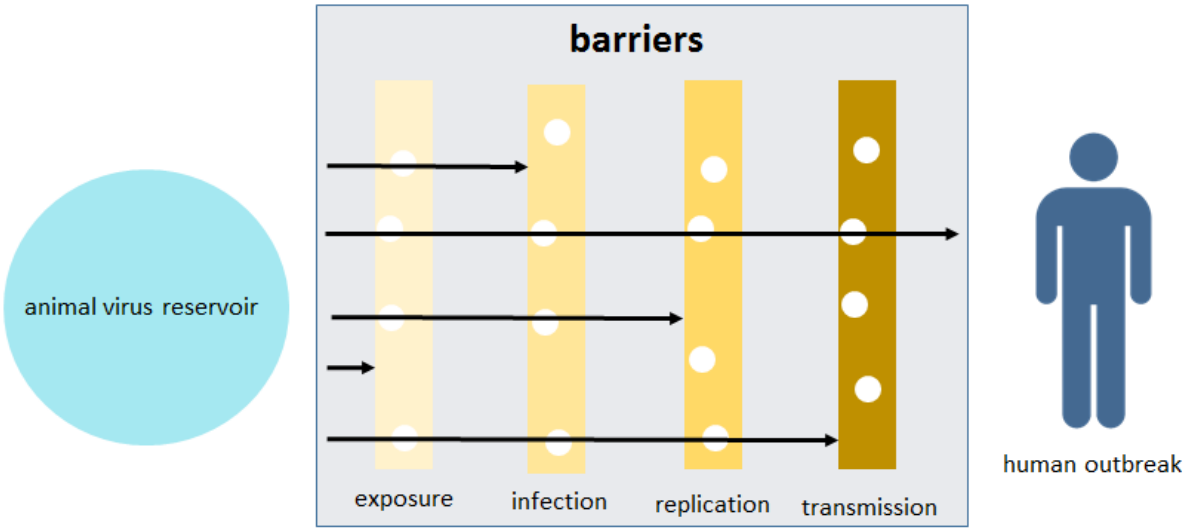


Figure 1 Barriers to endemic emergence.

In order to transmit from animals to humans a virus must penetrate a series of barriers. The holes represent gaps that allow the virus to overcome these barriers. An outbreak requires that gaps in all of the barriers are penetrated by the virus (adapted from (Plowright et al., 2017)). Note that this depiction is conceptual and does not reflect the total complexity of barriers involved in animal to human endemic emergence.

1.1.4. Influenza A virus (IAV)

1.1.4.1. The IAV virion and replication

The IAV virion is pleomorphic. This means that the structural protein, lipid and RNA components form intrinsic variable structures (Harris et al., 2006). Accordingly, the virions can be spherical, kidney-shaped or elongated, with average sizes of 120 nm as outer diameter (Harris et al., 2006). The virion is formed by the lipid envelope, RNA and proteins. Proteins that are part of the virion and participate in forming its structure are listed below (Figure 2):

- The spike proteins Hemagglutinin (HA), Neuraminidase (NA) and the proton channel (matrix protein 2) M2 that are embedded in the viral lipid envelope.
- The matrix protein 1 (M1) that resides on the inner side of the envelope and binds both, lipid envelope and the viral RNA.
- The Nucleoprotein (NP)
- The trimeric RNA dependent RNA polymerase (RdRp), which is composed of polymerase acidic protein (PA), polymerase basic protein 1 (PB1) and polymerase basic protein 2 (PB2).

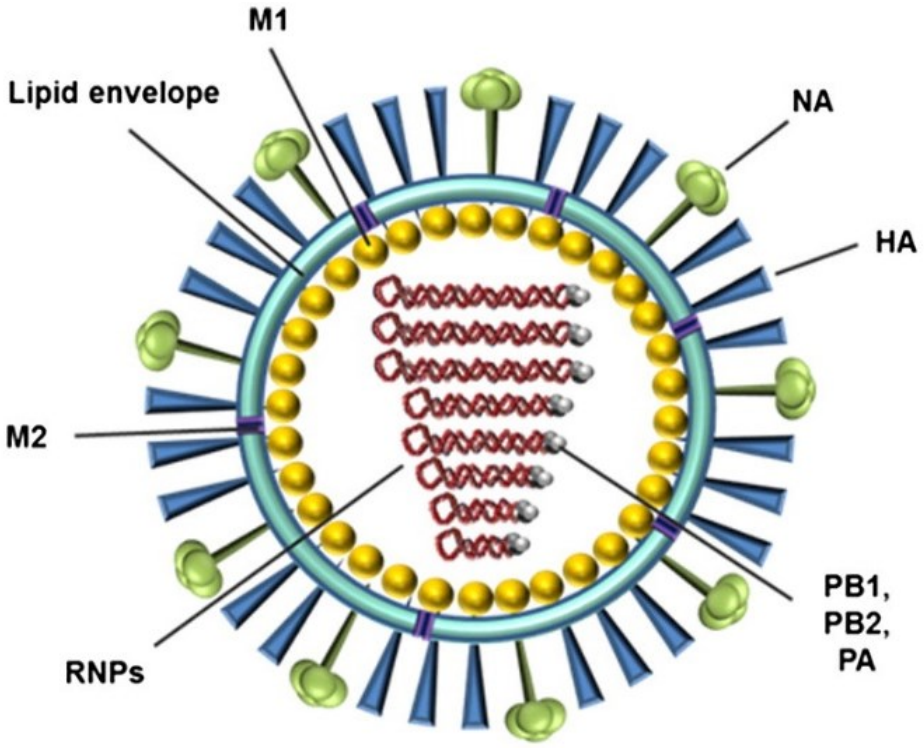


Figure 2 Schematic depiction of the IAV virion structure.

According to (Mair et al., 2014).

The viral genome is composed of eight single-stranded RNA segments of negative polarity, which are packaged with NP and the RdRp as viral ribonucleoproteins (vRNPs). In addition to the structural proteins listed above, the presence of the non-structural protein 1 (NS1) and 2 (NS2, also nuclear export protein NEP) in purified virions (Hutchinson et al., 2014) may indicate a role for these proteins in forming the structural integrity of the virion or regulation of post-entry infection steps.

The virion attaches to the cell by binding of the HA protein to sialic acid containing receptors that reside on the cell surface. The virus enters as an endosome after endocytosis (Figure 3). The acidic environment of the endosome triggers fusion of viral and endosomal membranes by inducing conformational changes to the HA (Helenius, 1992). The inner part of the endosome is acidified by proton influx via the

M2 ion channel. This leads to conformational changes to the M1 protein and its liberation from vRNPs (Martin and Helenius, 1991a, b). The vRNPs are then transported into the nucleus via recognition of nuclear localization signals (NLS) (Cros et al., 2005).

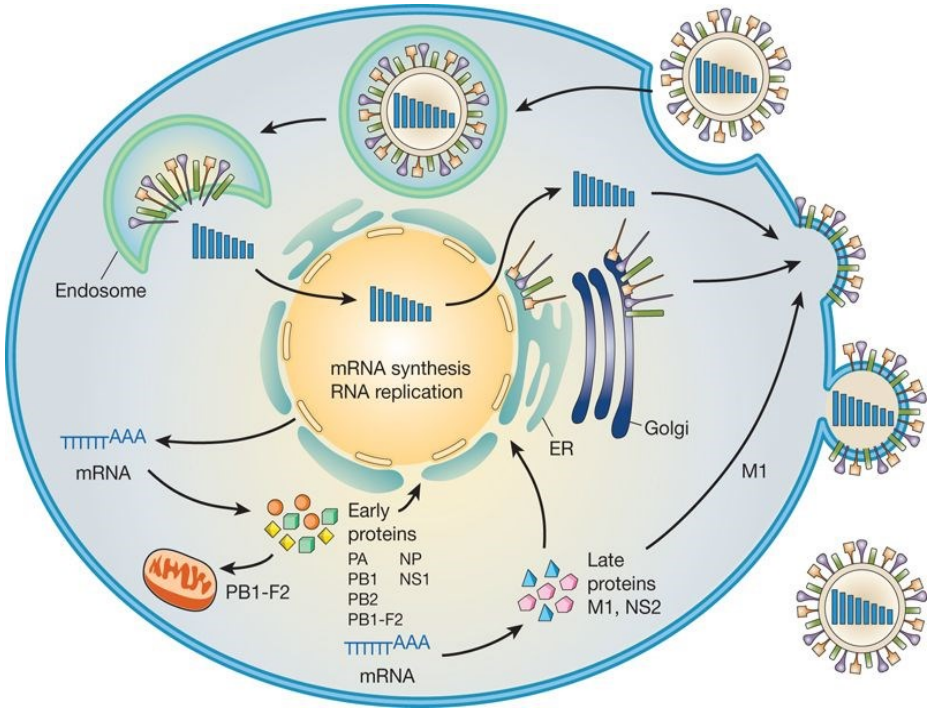


Figure 3 IAV replication cycle.

After attachment to the cells surface the virus enters the cell via endocytosis. vRNPs are liberated and transported to the nucleus, which is the site of replication and transcription. mRNAs are then translated in the cytoplasm. Viral proteins that assist in replication and transcription are imported into nucleus. PB1-F2 is targeted to mitochondria and membrane proteins are synthesized at the ER. M1 and NS2 mediate the export of vRNPs to the cytosol. Viral progeny bud from the plasma membrane (Neumann et al., 2009).

The nucleus is the site of transcription and replication of the viral genome. Both processes are catalyzed by the RdRp. Transcription of messenger RNAs (mRNAs) is initiated with capped RNA primers that are cleaved off from cellular mRNAs via the endonuclease activity of the PA protein (Dias et al., 2009; Plotch et al., 1981). The endonuclease activity contributes to downregulating the half-lives of cellular mRNAs, which allows the virus to take over the cells mRNA pool and consequently to produce

selectively viral proteins (Bercovich-Kinori et al., 2016; Jagger et al., 2012). The capped-primers serve to initiate transcription with the viral RNA as template (Plotch et al., 1981). Viral mRNAs are then poly-adenylated by transcription ("reiterative stuttering") of a stretch of uridines at the conserved 5' end of the viral RNA (Luo et al., 1991). In addition, two transcripts originating from the two smallest genome segments can be further spliced to alternative isoforms (see 1.1.4.2. IAV RNA splicing). The spike and membrane proteins of IAV are then translated at the endoplasmic reticulum (ER), processed in the *cis*-Golgi and transported via the *trans*-Golgi network to the membrane where they assist in budding of new viral progeny. The other viral proteins are synthesized in the cytoplasm and some are directed towards the nucleus where they assist in replication and transcription (Smith et al., 1987). In total, IAV encodes one major protein per segment, plus additional two proteins that are translated from spliced isoforms. Evidence has also been presented for the production of six additional accessory proteins (PB1-F2, PB1-N40, PA-X, PB2-S1, M42 and NS3) that may be produced by some viral isolates (Long et al., 2019).

Since the genetic material is of negative sense orientation, the polymerase needs to first transcribe the viral RNA into copy intermediates of positive sense orientation. These copy intermediates serve as templates for transcription into viral RNAs, which can be packaged into newly assembled virions.

Unlike transcription, that functions in *cis* (meaning that the polymerase sitting on the vRNP catalyzes transcription), replication of the viral genome functions in *trans* (meaning that a template vRNP is transcribed by a soluble polymerase other than the one that mediates encapsidation) (Eisfeld et al., 2015; Jorba et al., 2009).

The progeny vRNAs are complexed with NP and the RdRp and subsequently exported to the cytosol involving the cellular XPO1/CRM1 (Exportin-1/ Chromosome Region Maintenance 1 homolog) -dependent nuclear export pathway (Eisfeld et al., 2015). During this step the two viral proteins M1 (Bui et al., 2000; Martin and Helenius, 1991a) and NS2/NEP (Akarsu et al., 2003; O'Neill et al., 1998) play essential roles.

Then, all subviral structural components are directed to the plasma membrane (for polarized epithelial cells, the apical membrane) and are assembled in lipid rafts (nonionic detergent resistant lipid microdomains). M1 is thought to bridge vRNPs to the cytosolic tails of HA and NA proteins, thus facilitating the close spatial assembly of these structural components (Nayak et al., 2004). Progeny virions are sialylated (covalent addition of sialic acid moieties) by intracellular enzymes and aggregate to each other due to the interaction of HA with sialic acids (Wagner et al., 2002). The enzymatic activity of NA is essential for buddying as it destroys sialic acid receptors from the viral envelope to avoid the aggregation of progeny virions (Palese et al., 1974).

1.1.4.2. IAV RNA splicing

IAV uses the cellular splicing machinery to process RNAs from the two smallest genomic segments (segments 7 and 8, also referred to as M segment or NS segment, respectively) (Dubois et al., 2014). The M segment was shown to encode for 4 RNAs (Figure 4a): One collinear transcript (M1 mRNA) that can be translated to the M1 protein, one alternatively spliced M2 mRNA that can be translated to the M2 ion channel (Lamb et al., 1981), an RNA 3 (Shih et al., 1998), which is not known to encode a peptide and the M4 mRNA, which is proposed to be translated to an isoform of the M2 ion channel in certain strains (Wise et al., 2012). The NS segment was shown to encode three RNAs (Figure 4b): One collinear transcript (NS1 mRNA) that can be translated to the NS1 protein, an alternatively spliced product that encodes the NS2/NEP protein (Lamb and Choppin, 1979) and an alternatively spliced NS3 mRNA that is produced by some strains and is linked to host switching (Selman et al., 2012). Importantly, the splicing events that have been described for each of the segments use different 5' donor sites but the same 3' splice site. Several *cis*-acting signals (i.e. RNA secondary structures (Ferhadian et al., 2018) and splicing enhancer motifs (Huang et al., 2017)) and *trans*-acting factors (i.e. NS1 (Robb and Fodor, 2012), HNRNPK (Heterogeneous Nuclear Ribonucleoprotein K) (Tsai et al., 2013), NS1-BP (NS1-binding protein) (Thompson et al., 2018), ASF/SF2 (Huang et al., 2017; Shih and Krug, 1996)) were found to play a role in M/NS segment RNA splicing.

The timely production due to regulated splicing is critical for a permissive infection. For example, the ratio of spliced to unspliced products from the M segment increases during infection, which reflects the changing demands for optimal viral replication (Valcarcel et al., 1991). Also, NS splicing is regulated in a way that NS2/NEP is

produced at a time when it is needed to promote the export of vRNPs from the nucleus (Chua et al., 2013).

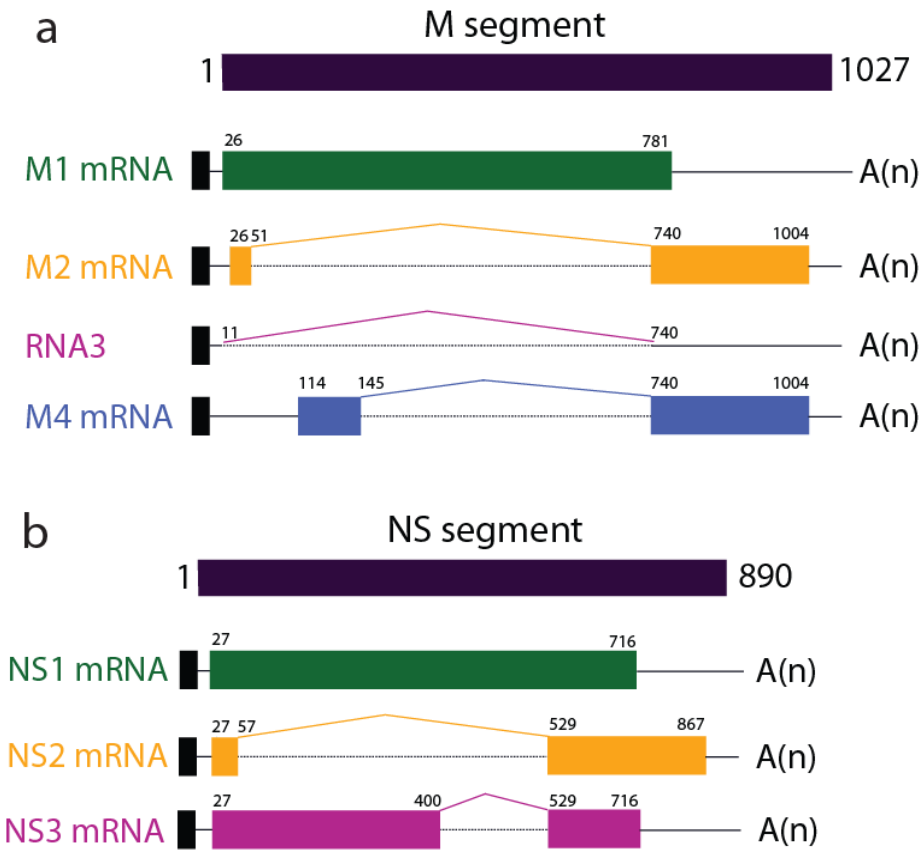


Figure 4 Schematic depiction of M and NS gene architectures.

*Transcripts from the M **(a)** and NS **(b)** segments are either collinear (green) or can further be spliced into the indicated isoforms (yellow, purple and blue, respectively). Open reading frames (ORFs) are highlighted by solid boxes. The numbers above indicate the positions of the first and last nucleotides in the ORF as well as the positions of splice donor and acceptor sites.*

1.1.4.3. IAV epidemics and pandemics

Influenza A viruses (IAVs) are causing an acute respiratory disease (called the “flu”) that includes symptoms such as fever, headache, runny nose, sore throat, etc. Influenza occurs seasonally in epidemic waves with peak infection rates during the winter season in the northern hemisphere. It is estimated that Influenza results in about 3 to 5 million cases of severe illness, and about 290,000 to 650,000 respiratory deaths annually (Iuliano et al., 2018; World Health Organization, 2019a). In addition to seasonal epidemics, IAVs occasionally cause pandemic outbreaks. This situation occurs when strains that are very different from circulating seasonal IAVs acquire the ability to transmit among humans (Long et al., 2019; Neumann and Kawaoka, 2015). Table 1 gives an overview on the pandemic outbreaks that were recorded during the 20th and 21st century.

Table 1 Summary of the pandemics caused by IAV in the 20th and 21st century.

Table according to (Saunders-Hastings and Krewski, 2016).

Year	Pandemic name	Strain	Approximate number of deaths
1918-1920	Spanish flu	H1N1	40-50 million
1957-1958	Asian flu	H2N2	1-2 million
1968-1970	Hong Kong flu	H3N2	500,000-2 million
2009-2010	Swine flu	H1N1	Up to 575,000

Notably, IAV pandemics are associated with high mortalities, which makes understanding of the underlying processes a top priority for infection research. During a pandemic wave a large proportion of the population is infected and subsequently mounts adaptive immunity towards the specific strain. Strains that were introduced during a pandemic wave reappear later as seasonal epidemics, when gradual

alterations to the HA and NA antigens were introduced (Taubenberger and Kash, 2010).

The three most recent pandemics were linked to reassortment events (Saunders-Hastings and Krewski, 2016; Worobey et al., 2014a) (that is, the exchange of gene segments during co-infection with different IAV strains). For example, co-infection of a cell with an avian Influenza virus (AIV) and a human virus can result in the acquisition of human-adapted polymerase gene segments by a virus with novel AIV-derived antigens to which the human population is immunologically naïve (Long et al., 2019). In such a manner the virus of the Asian Flu gained PB1, HA and NA segments from duck and the other segments from human IAV. In addition to the rapid adaptation, known as antigenic shift, gradual adaptation of avian viruses to humans is also a plausible pathway for pandemic emergence (Long et al., 2019). This concern is raised due to numerous cases where humans were infected with avian strains to which humans are immunological naïve, such as H7N9 (Cowling et al., 2013), H5N1 (Tran et al., 2004), H9N2 (Butt et al., 2005), H10N8 (World Health Organization, 2019b). Most of these infections were dead ends, meaning that the strains could not acquire sustained human to human transmission and occurred only in a low number of individuals. However, due to their potential to cause pandemic outbreaks, ongoing infections with these strains are under constant surveillance by the WHO (world health organization, (World Health Organization, 2019b)).

1.1.4.4. Avian IAV

Although IAV can infect a wide variety of hosts, the natural host of the virus is wild waterfowl (e.g. ducks, mallards, geese, seagulls, etc.). Subtypes that infect wild waterfowl are spread globally. Due to migration of the birds Influenza virus subtypes and strains are constantly mixed and spread to domestic poultry. In contrast to humans and mammals, IAVs replicate in cells of the intestinal tract in birds. Viruses are then shed in high concentration as part of the feces into surface waters (Olsen et al., 2006).

On the basis of their pathogenicity in birds, avian IAV can be categorized as either low pathogenic (LPAI – low pathogenic avian Influenza) or high pathogenic (HPAI – high pathogenic avian Influenza). While the majority of infections in wild waterfowl is asymptomatic or subclinical, specific subtypes of HPAI-type cause occasional outbreaks in domestic poultry with high morbidity. These highly-virulent strains were initially termed “fowl-plague-virus”.

The major functional determinant of pathogenicity in poultry is the cleavage site in the HA protein (Chen et al., 1998). The HA that is expressed at the cell surface must be cleaved in order to be functional (that is, promote membrane fusion once the released virions enter new cells). The HA of LPAI isolates is cleaved at a single arginine residue (*monobasic cleavage site*) by trypsin-like proteases that are expressed in the intestinal tract of birds. In contrast, the HA of HPAI isolates can be cleaved at multiple residues (*polybasic cleavage site*) by furin-like proteases that are more ubiquitously expressed. This allows HPAI to spread systemically in chicken and induce more severe disease (Mair et al., 2014).

1.1.4.5. Molecular determinants of species specificity

The species barriers that hinder most avian IAVs from successfully infecting humans are effective at several steps in the viral life cycle.

The first barrier represents the entry of Influenza A virions in the cell. This happens via binding of HA to sialic acid containing receptors. The sialic acid moieties on cell surface proteins are linked to galactose by α-2,6 or α-2,3. The dominant linkage in the human upper respiratory airway epithelium is of α-2,6 type, to which human adapted HA binds preferentially (Matrosovich et al., 1997). Despite these differences in receptor binding, many avian viruses are internalized by human cells and initiate expression of the viral genome. Such infections typically lead to an abortive, nonproductive outcome in human cell lines (Inglis and Brown, 1984; Israel, 1980; Lau and Scholtissek, 1995; Sadewasser et al., 2017). Our understanding of this intracellular restriction is still incomplete. However, a few important determinants that were addressed in previous studies are listed below:

- After endocytosis, fusion of viral and endosome membrane occurs. This process requires conformational changes in the HA protein that depend on the pH of the endosome (Bullough et al., 1994) and activity of cellular proteases (Mair et al., 2014). Interestingly, HA from human isolates tend to be more stable at low pH than HAs from avian isolates (Galloway et al., 2013).
- After liberation of the vRNPs, vRNPs are detected by host sensors such as RIG-I (retinoic acid-inducible gene I) (Rehwinkel et al., 2010), which stimulates the interferon response. Interestingly, RIG-I has also a second antiviral effector function during IAV infection. It detects vRNPs and inhibits the polymerase

activity. Importantly, avian-adapted strains are more sensitive to inhibition by RIG-I, which is at least partially due to a one amino acid difference at amino acid position 627 in PB2 that differs between avian and human-adapted IAV strains (Weber et al., 2015).

- Avian IAV strains are also impaired in their ability to import essential viral proteins, such as PB2 and NP, or vRNPs to the nucleus. For example, PB2 proteins from mammalian-adapted strains bind importin-α stronger than that of bird-adapted isolates (Resa-Infante et al., 2008). Also, the specificity as to which importins are used for nuclear entry differs between avian and mammalian adapted IAV strains. NP and PB2 of avian viruses use preferentially importin-α7, while human-adapted strains prefer importin-α3 (Gabriel et al., 2011).
- One important and well-established factor that contributes to host range is the viral RdRp. Polymerases from avian strains are considerably less active in mammalian cells than their counterparts from mammalian-adapted strains (Gabriel et al., 2005; Long et al., 2016). The protein ANP32A (Acidic Nuclear Phosphoprotein 32 Family Member A) was found to be a critical factor controlling host-specific activity of the RdRp (Long et al., 2016). In addition, a restriction factor of unknown identity may exist that is more sensitive to avian than human polymerases (Mehle and Doudna, 2008).
- An important restriction factor that targets vRNPs is the cytosolic interferon inducible dynamin-like GTPase MxA (Staeheli et al., 1986). Upon infection MxA recognizes vRNPs and oligomerizes, thus inhibiting the functions of vRNPs

(Gao et al., 2011). Interestingly, vRNPs from avian-adapted viruses are more sensitive to inhibition by MxA than vRNPs of human origin (Dittmann et al., 2008). The adaptive mutations controlling this sensitivity were mapped to the NP protein (Manz et al., 2013).

1.2. Proteomics

1.2.1. Quantitative shotgun-proteomics

Proteins are a very diverse class of biological macromolecules that facilitate all processes during a viral infection. Studying proteins is thus of paramount interest when investigating these processes. Traditionally, this is achieved by targeted assessment of individual proteins, such as by antibody-based (ELISA, Immunoblotting, Immunoprecipitation, etc.) assays or by genetically modifying individual proteins (e.g. inserting tags).

The advent of mass-spectrometry (MS) based proteomics has enabled to study many proteins in parallel (that is, the proteome). A major breakthrough in mass-spectrometry based proteomics was the development of soft ionization techniques (ESI and MALDI) that allow biological macromolecules to be analyzed by MS. Many improvements in instrumentation (such as development of the orbitrap mass analyzers (Makarov et al., 2006) or fragmentation techniques), software (e.g. for peptide and protein identification and quantification (Cox and Mann, 2008; Cox et al., 2011)), and experimental procedures (Rappsilber et al., 2003) made it feasible to routinely study thousands of proteins in parallel in single experiments (Nilsson et al., 2010).

A typical shotgun proteomics experiment involves the digestion of proteins to peptides using sequence-specific proteases. The peptide mixture is then separated via e.g. reversed-phase liquid chromatography on columns with microscale diameter that are coupled to a mass-spectrometer. The m/z (mass divided by charge) ratio of the analyte and its ion count at each m/z value is recorded in the mass analyzer (Aebersold and Mann, 2003). To identify the sequence of a peptide, the mass-

spectrometer obtains its mass and its fragmentation pattern in an approach called tandem mass spectrometry. Algorithms such as Andromeda (Cox et al., 2011) then identify peptides by matching these data to protein databases. Approaches such as labeling with tandem mass tags (TMT) or SILAC then allow to accurately quantify differences in protein abundance between samples (Ong et al., 2002; Thompson et al., 2003).

The versatility and power of the approach make it an ideal tool to study virus infection. Accordingly, quantitative shotgun proteomics contributed to many important discoveries in virology (Greco et al., 2014).

1.2.2. Assessment of proteome dynamics

Many processes during a viral infection happen during a defined stage of infection. To analyze these processes it is critical to specifically label proteins that are synthesized within a narrow time window. One possibility how mass-spectrometry-based analysis can uncover synthesis and turnover of proteins is by labeling with SILAC during defined pulse intervals (pulsed SILAC, pSILAC) (Schwanhausser et al., 2009). The depth (that is, the number of proteins identified and quantified) of pSILAC experiment is limited due to the fact that proteins that incorporated SILAC amino acids are analyzed in a mixture with proteins that did not. The use of bioorthogonal amino acids containing reactive groups has facilitated the analysis of proteome dynamics, since the reactive side chains can be used to specifically enrich proteins synthesized during a specific time window.

For example, the molecule azidohomoalanine (AHA) shares chemical similarity to methionine and accordingly AHA (Figure 5) can be incorporated into newly synthesized proteins, using the endogenous Met t-RNA synthetase (Dieterich et al., 2006). The newly synthesized proteins can then be covalently coupled to alkyne groups using a copper-catalyzed cycloaddition in aqueous solution (Dieterich et al., 2006). AHA has now been successfully used to study secreted proteins (Eichelbaum and Krijgsveld, 2014a), the changes in protein synthesis upon macrophage stimulation (Eichelbaum and Krijgsveld, 2014b) and protein degradation (McShane et al., 2016). It shows minimal toxicity, can be administered to mice and is incorporated into various tissues (Calve et al., 2016). Also, AHA has been used in combination with SILAC to quantify proteome dynamics (Eichelbaum and Krijgsveld, 2014a, b; McShane et al., 2016).

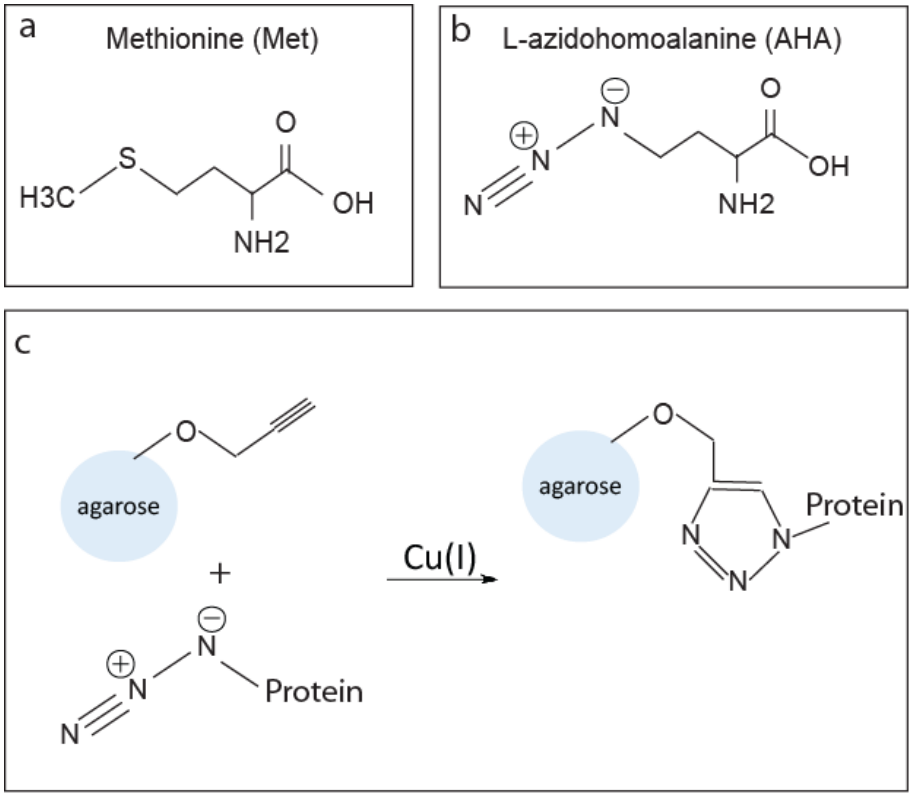


Figure 5 AHA allows the selective enrichment of newly synthesized proteins.

*The methionine **(a)** analogous amino acid L-azidohomoalanine **(b)** is integrated into newly synthesized proteins when medium is methionine depleted. Newly synthesized proteins can be selectively coupled to alkyne containing reagents (e.g. alkyne agarose beads) in a copper catalyzed click-chemistry reaction **(c)**.*

1.2.3. Affinity purification-mass spectrometry in virology

Proteins typically do not act alone, but require the action of binding partners to exert their function. Accordingly, studying protein-protein interactions is central to understanding host-virus relationships.

There are many experimental system to study protein protein interactions. Some, such as FRET (Förster resonance energy transfer) or ITC (isothermal titration calorimetry), can give accurate results about binding affinities. Others, such as NMR (nuclear magnetic resonance), give structural insights into binding interfaces. However, these methods require prior knowledge about the involved interaction partners (Meyer and Selbach, 2015).

Over the recent years, affinity purification-mass spectrometry (AP-MS) experiments have become the gold standard for globally studying host virus interactions (Batra et al., 2018; Jager et al., 2011; Pichlmair et al., 2012; Ramage et al., 2015; Rozenblatt-Rosen et al., 2012; Scaturro et al., 2018; Shah et al., 2018). The principle is based on transfecting expression vectors encoding for tagged baits into cells and then affinity enrich the bait and its interaction partners and analyzing the samples by mass-spectrometry. These approaches have delivered important insights as to the interaction networks of proteins from different viruses such as HCV (Ramage et al., 2015), HIV (Jager et al., 2011), Flaviviruses (Shah et al., 2018), Zika virus (Scaturro et al., 2018), Ebola virus (Batra et al., 2018) or tumor viruses (Rozenblatt-Rosen et al., 2012). AP-MS approaches were also used to compare interactors of proteins from different viruses that are functionally similar (e.g. viral immune modulators (Pichlmair et al., 2012)).

However, it is imperative to discriminate background binders from true interaction partners in AP-MS experiments. This is typically achieved by performing control purifications from samples that do not express the tagged bait. In these cases the use of labeling strategies (such as SILAC or TMT labeling) can be beneficial as it allows the researcher to multiplex samples from the immunoprecipitation with control samples.

2. AIMS AND OBJECTIVES

Influenza A viruses are of great importance to human health (see 1.1.4.3. IAV epidemics and pandemics). One important question is how pandemic outbreaks arise. From the genesis of these outbreaks it is clear, that avian viruses gained the ability to replicate in humans and be transmitted between humans. While some aspects of the adaptation of avian viruses to a human host were addressed previously (see 1.1.4.5. Molecular determinants of species specificity), the problem is yet poorly understood.

One way to assess the adaptive mechanisms that are involved in crossing the species barrier is by investigating human and avian IAVs in cell culture. Such a model system was developed by the laboratory of Thorsten Wolff (Robert-Koch-Institute, Berlin). Specifically, a low-pathogenic avian H3N2 IAV (A/Mallard/439/2004 – “Mal”) was compared to a seasonal human IAV isolate of the same subtype (A/Panama/2007/1999 – “Pan”). The avian-adapted virus does not multiply efficiently in cultured human cells and causes a non-permissive infection. In contrast, the seasonal human-adapted virus produces a high amount of viral progeny. At the Robert-Koch-Institute, it was demonstrated that the Pan virus produces >1,000 fold more infectious viral progeny than the non-adapted virus. Both strains enter efficiently human cells and start viral gene expression program (Sadewasser et al., 2017) (Figure 6).

We hypothesized that this restriction must occur intracellularly and that analysis of proteome dynamics can shed light on the host and viral determinants of permissiveness. First, it was aimed to use the methionine analogue amino acid AHA

to compare the synthesis of host and viral proteins under infection with the non-adapted and the adapted strain at a global scale. Second, we aimed to integrate our proteome level data with mRNA level data to understand to what extent mRNA level changes explain the levels of synthesized proteins. Third, we specifically analyzed the gene expression regulation of the M1 gene, an essential viral gene that was expressed with the highest difference between both strains. It was assessed whether differential splicing contributes to the differences in M1 expression levels between both strains. Also, the sequence determinants that control strain-specific splicing should be determined. Lastly, it was addressed to what extent differences in the regulation of M segment RNA splicing contribute to species specificity of IAV.

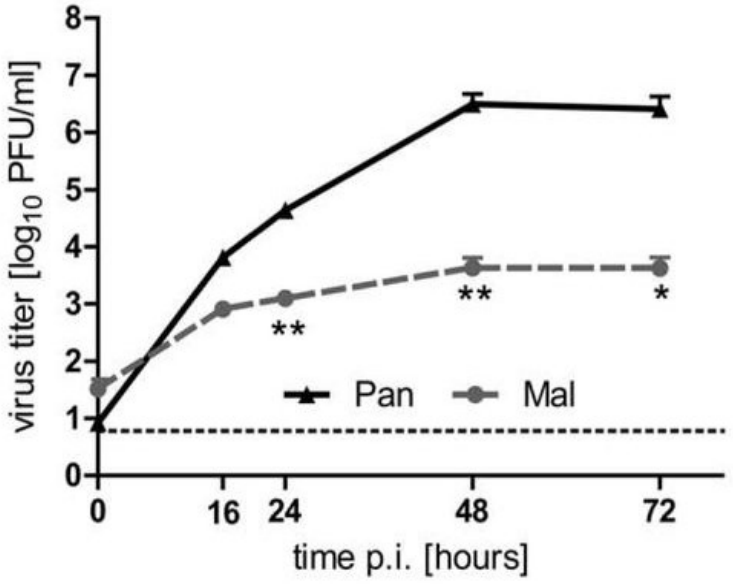


Figure 6 Comparison of an avian and a human adapted IAV isolate.

Multicycle replication curves of Pan and Mal viruses on human lung epithelial cells (A549) infected with a low MOI = 0.01. Supernatants were harvested at the indicated time points and subjected to titration on MDCKII cells. Figure courtesy of Anne Sadewasser (Sadewasser et al., 2017)

3. MATERIALS

3.1. Chemicals

All chemicals were obtained from Carl Roth (Karlsruhe, Germany), Merck (Darmstadt, Germany), Life Technologies (Carlsbad, USA) or Sigma-Aldrich (Steinheim, Germany), unless stated otherwise.

3.2. Cell lines

- A549

Human lung epithelial adenocarcinoma cells obtained from ATCC. A549 cells are a model system for IAV infections in humans.

- MDCK type II

Madin-Darby Canine Kidney cells obtained from ATCC. MDCK type II cells are highly permissive for IAV infections.

- DF-1

Chicken fibroblast cells obtained from ATCC. Avian cells that are permissive for avian IAV infection.

- HEK293T

Human embryonic kidney cells expressing SV40 large T antigen. A cell line that can be easily transfected.

All cell lines were routinely checked for Mycoplasma contamination.

3.3. Virus strains

The following virus strains were used for infection/transfection experiments:

- A/Mallard/439/2004 (Mal)

A low-pathogenic avian IAV strain of H3N2 subtype isolated from Mallard duck. The genome sequence is available under GISAID accession numbers: EPI859640-EPI859647.

- A/Panama/2007/1999 (Pan)

Pan is a seasonal human IAV isolate of H3N2 subtype. The genome sequence is available under the NCBI accession numbers: DQ487333-DQ487340.

- Pan + Mal M

A 7+1 reassortant strain carrying seven Pan genomic segments, plus the M segment of the Mal strain.

- Pan-AV

A mutant virus where the following positions at the M segment 3' splice site region were changed to the respective sequence of the Mal strain: A712G, G714A, A740G, T745C, A754G, G760A, G766A, A772G.

In addition M segments were cloned from the following avian, human or swine IAVs:

- A/chicken/Rostock/45/1934 (FPV)

An HPAI isolate of H7N1 subtype. One of the earliest avian IAV isolates of the eastern avian IAV lineage. The M segment sequence is available under the NCBI accession: CY077423

- A/Vietnam/1203/2004

A zoonotic IAV isolate of H5N1 subtype that caused a fatal case in humans (Tran et al., 2004). The M segment sequence is available under the accession HM006762.1

- A/swine/Netherlands/25/1980

An isolate of IAV obtained from swine of H1N1 subtype. It is an early isolate of the “avian-like” swine lineage (Smith et al., 2009). The M segment sequence is available under the accession: Z26862.1.

- A/BrevigMission/1/1918

The pandemic isolate of IAV that caused the “Spanish Flu” outbreak. This virus is of H1N1 subtype and the M segment sequence is available under the accession: AY130766

- A/WSN/1933

This virus originated from the first isolation of IAV in humans (Smith et al., 1995). This strain was then passaged serially in mice brain, giving rise to a virus with neurotropic character (Francis and Moore, 1940). In this thesis, NP, PB1, PA, and PB2 segments of this strain were used for the RdRp reporter system. The segment sequences are available under the following NCBI accessions: LC333182-LC333189.

3.4. Plasmids

The plasmids that were used in the present thesis are listed in Table 2.

Table 2 Plasmids used in this thesis.

IDs 6-9 and 11-14 were created in collaboration with Barbara Vetter, Charité, Berlin. IDs 16-21 were created in collaboration with Martha Hergeselle, MDC, Berlin.

ID	Plasmid	Source
1	pDONR221	ThermoFisher
2	pDEST26-Flag/HA	Katrina Meyer, MDC Berlin
3	pDEST26-Flag/HA-NS1(Mal)	this thesis
4	pDEST26-Flag/HA-NS1(Pan)	this thesis
5	pDEST26-Flag/HA-Mseg(Mal)29-1007	this thesis
6	pDEST26-Flag/HA-Mseg(Mal)29-1007 w/ Pan 326-706	this thesis
7	pDEST26-Flag/HA-Mseg(Mal)29-1007 w/ Pan 707-779	this thesis
8	pDEST26-Flag/HA-Mseg(Mal)29-1007 w/ Pan 277-779	this thesis
9	pDEST26-Flag/HA-Mseg(Mal)29-1007 w/ Pan 52-739	this thesis
10	pDEST26-Flag/HA-Mseg(Pan)29-1007	this thesis
11	pDEST26-Flag/HA-Mseg(Pan)29-1007 w/ Mal 326-706	this thesis
12	pDEST26-Flag/HA-Mseg(Pan)29-1007 w/ Mal 707-779	this thesis
13	pDEST26-Flag/HA-Mseg(Pan)29-1007 w/ Mal 277-779	this thesis
14	pDEST26-Flag/HA-Mseg(Pan)29-1007 w/ Mal 52-739	this thesis
15	pDEST26-Flag/HA-M2	this thesis
16	pDEST26-Flag/HA-Mseg(FPV)	this thesis
17	pDEST26-Flag/HA-Mseg(p1918)	this thesis
18	pDEST26-Flag/HA-Mseg(A/Vietnam/1203/2004)	this thesis
19	pDEST26-Flag/HA Mseg(A/swine/Netherlands/25/1980)	this thesis
20	pDEST26-Flag/HA Mseg(p1918) w/ Mal 707-779	this thesis
21	pDEST26-Flag/HA Mseg(FPV) w/ Pan 707-779	this thesis
22	pDEST26-cmyc	Marieluise Kirchner, MDC Berlin
23	pDEST26-cmyc-NS1(Mal)	this thesis
24	pDEST26-cmyc-NS1(Pan)	this thesis
25	pHW2000-Mal M	Katharina Paki, RKI Berlin
26	pHW2000-Pan-Av	this thesis
27	pHW2000-Mal-Hu	this thesis
28	pHW2000-Pan (segment1-8)	RKI
29	pHW2000-WSN(segment1-3, 6)	RKI
30	pDEST-EGFP	Markus Landthaler, MDC

3.5. Buffers and solutions

3.5.1. Virus infection and cell culture

- DMEM culturing medium 1

10 % fetal bovine serum (FBS) [v/v], 2 mM glutamine (Glutamax, Life Technologies), 50 mg/mL penicillin/ streptomycin (Life Technologies), Dulbecco's modified Eagle medium (DMEM, Life Technologies)

- DMEM culturing medium 2

10 % FBS [v/v], 4 mM glutamine, 50 mg/mL penicillin/ streptomycin, DMEM

- MEM culturing medium

10 % FBS [v/v], 2 mM glutamine, 50 mg/mL penicillin/ streptomycin, minimum essential medium Eagle (MEM, Merck)

- SILAC DMEM light / medium / heavy

10 % dialyzed FBS (Pan-Biotech), SILAC DMEM (Gibco), 2 mM glutamine, 50 mg/mL penicillin/ streptomycin, 84 mg/mL light L-arginine (Arg0), 146 mg/mL light L-lysine (Lys0) (for medium heavy SILAC DMEM: 84 mg/mL medium L-arginine (Arg6), 146 mg/mL medium L-lysine (Lys4) ; for heavy SILAC DMEM: 84 mg/mL heavy L-arginine (Arg10), 146 mg/mL heavy L-lysine (Lys8))

- Infection SILAC DMEM light / medium / heavy

SILAC DMEM (Gibco), 2 mM glutamine, 50 mg/mL penicillin/ streptomycin, 84 mg/mL light L-arginine (Arg0), 146 mg/mL light L-lysine (Lys0) (for medium heavy SILAC DMEM: 84 mg/mL medium L-arginine (Arg6), 146 mg/mL medium L-lysine (Lys4) ; for

heavy SILAC DMEM: 84 mg/mL heavy L-arginine (Arg10), 146 mg/mL heavy L-lysine (Lys8))

- Infection DMEM

0.2 % [w/v] bovine serum albumin (BSA, PAA Laboratories), 2 mM glutamine, DMEM

- pulsed SILAC Infection DMEM medium / heavy

2 mM glutamine, SILAC DMEM, 84 mg/mL medium L-arginine (Arg6), 146 mg/mL medium L-lysine (Lys4) (for heavy SILAC DMEM: 84 mg/mL heavy L-arginine (Arg10), 146 mg/mL heavy L-lysine (Lys8))

- Infection SILAC DMEM light / medium / heavy w/o methionine

SILAC DMEM w/o methionine, lysine and arginine (Biosera Ltd), 2 mM glutamine, 50 mg/mL penicillin/ streptomycin, 84 mg/mL light L-arginine (Arg0), 146 mg/mL light L-lysine (Lys0) (for medium heavy SILAC DMEM: 84 mg/mL medium L-arginine (Arg6), 146 mg/mL medium L-lysine (Lys4) ; for heavy SILAC DMEM: 84 mg/mL heavy L-arginine (Arg10), 146 mg/mL heavy L-lysine (Lys8))

- Phosphate buffered saline (PBS)

137 mM sodium chloride, 2.7 mM potassium chloride, 80.9 mM sodium hydrogen phosphate, 1.5 mM potassium hydrogen phosphate

- $PBS^{+/+}$

PBS, 0.2 % [w/v] BSA, 0.1 g/L magnesium chloride, 0.13 g/L calcium chloride

3.5.2. Bacterial microbiology

- LB-medium

Pre-mixed LB (lysogeny broth) powder (Sigma-Aldrich), 20 g/L dissolved in distilled water and autoclaved

- LB-agar

Pre-mixed LB (lysogeny broth) with Agar (Sigma-Aldrich), 35 g/L dissolved in distilled water, autoclaved, supplemented with antibiotics (see below) and poured in petridishes

- Ampicillin

100 mg/mL in distilled water, sterile filtered, diluted 1:1,000 for use

- Kanamycin

50 mg/mL in distilled water, sterile filtered, diluted 1:1,000 for use

3.5.3. Cloning, mutagenesis and *in vitro* transcription

- TAE (tris-acetate-EDTA) buffer

40 mM Tris-HCl ph 8.5, 40 mM acetic acid, 1 mM ethylenediaminetraacetic acid (EDTA) in distilled water

- Native agarose gel

1-2 % [w/v] agarose in TAE buffer

- RNA refolding buffer

10 M Tris, 100 mM KCl, 10 mM $MgCl_2$ in nuclease-free water

- 10X MOPS buffer

0.2 M 3-(*N*-morpholino)propanesulfonic acid (MOPS), 50 mM sodium acetate, 10 mM EDTA, pH 7

- RNA denaturing sample buffer

65 % [v/v] formamide, 22 % [v/v] formalin (37 % [v/v] formaldehyde), 13 % [v/v] 10X MOPS buffer

- RNA loading dye

50 % [v/v] glycerol, 1 mM EDTA, 0.3 % [w/v] bromophenol blue

- Denaturing agarose gel

3 % [v/v] formalin (37 % [v/v] formaldehyde), 1X MOPS in nuclease-free water, 1 % [w/v] agarose

3.5.4. Transfection and extraction of eukaryotic cells

- Lysis buffer

125 mM NaCl, 0.1 % [w/v] SDS, 1 % [v/v] NP-40, 5 % [v/v] glycerol, 50 mM Tris-HCl, adjusted to pH 7.4, supplemented with 1 tablet / 10 mL protease inhibitor (Roche cOmplete Tablets)

- PEI solution

1 g/L [w/v] PEI (Polyethylenimine, Polysciences, Warrington, USA) in ddH_2O

3.5.5. LC-MS sample preparation

All solutions destined for use with LC-MS applications were prepared with LC-MS grade (LiChrosolv) water or acetonitrile.

- ABC (ammonium bicarbonate) buffer

50 mM ammonium bicarbonate in water

- Denaturation buffer

6 M Urea, 2 M thiourea in 10 mM HEPES, pH 8.0

- TFA

10 % [v/v] trifluoracetic acid in water

- Alkylation buffer

55 mM iodoacetamide in ABC

3.5.6. Immunofluorescence microscopy

- Fixation solution

2.5 % [v/v] formaldehyde in PBS

- Permeabilization solution

0.2 % [v/v] Triton-X-100 in PBS

- Staining solution 1

3 % BSA [w/v] in PBS w/ 1:1,000 NP-antibody

- Staining solution 2

3 % BSA [w/v] in PBS w/ 1:1,000 Alexa-Fluor 488 goat anti mouse secondary antibody containing 1 µg/ mL DAPI (4',6-diamidino-2-phenylindole)

- Mowiol solution

20 % [w/v] Mowiol 4-88, 50 % [v/v] glycerol in ddH_2O

3.5.7. SDS-PAGE and western blotting

- TBST (Tris-bufferd saline Tween-20) buffer

25 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05 % [v/v] Tween-20 in distilled water

- Blocking solution

6 % [w/v] skim milk powder in Tris-buffered saline Tween-20 (TBST)

- Transfer buffer

25 mM Tris-HCl pH 8.5, 190 mM glycine, 20 % [v/v] methanol in distilled water

3.5.8. Kits

- µMACS c-myc isolation kit (Miltenyi Biotec)
- µMACS HA isolation kit (Miltenyi Biotec)
- Click-iT Protein Enrichment Kit (Invitrogen), contains urea lysis buffer, alkyne resin, reaction additive 1, 100 mM copper-II-sulfate, SDS wash buffer and reaction additive 2
- Gibson Assembly cloning kit (New England Biolabs, Ipswich, USA)
- Invisorb Spin DNA extraction kit (STRATEC Molecular, Berlin, Germany)
- Invisorb Spin Plasmid Mini Two plasmid DNA purification kit (STRATEC Molecular)
- MegaScript kit for *in vitro* transcription (Life Technologies)
- Novex ECL HRP chemiluminescent substrate reagent kit (invitrogen, Carlsbad, USA)
- NucleoBond Xtra Midi plasmid DNA purification kit (MACHEREY-NAGEL, Düren, Germany)
- Protein concentration assay DC Protein Assay (BioRad, Munich Germany)

3.5.9. Antibodies

The following antibodies were used to specifically detect antigens:

- HA epitope (rat, clone 3F10, Roche), diluted 1:1,000 for western blotting
- VEZT (mouse, clone B-1, SantaCruz), diluted 1:1,000 for western blotting
- M1 (mouse, clone GA2B, Bio-Rad), diluted 1:500 for western blotting

- M2 (rabbit, polyclonal, RRID: AB_2549706, Thermo Fisher), diluted 1:1,000 for western blotting
- NP (mouse, clone AA5H, Serotec), diluted 1:1,000 for both immune fluorescence and FACS staining
- RUVBL1 (mouse, clone A4, SantaCruz), diluted 1:1,000 for western blotting
- Anti-rabbit, anti-mouse or anti-rat horseradish peroxidase-linked secondary antibodies (1:10,000, for western blotting, Sigma-Aldrich)
- Alexa Fluor 488-conjugated goat anti-mouse IgG antibody (1:1,000 for flow cytometry and immune fluorescence, Invitrogen)

3.5.10. Consumables

- 4 %–12 % Bis-Tris gradient gels (NuPAGE, Invitrogen)
- 500 mM Azidohomoalanine (Anaspec, Inc., Fremont, USA) dissolved in dimethylsulfoxide (DMSO)
- Gateway BP Clonase II enzyme mix (Life Technologies)
- Gateway LR Clonase II enzyme mix (Life Technologies)
- Gel Loading Dye, Purple 6X (New England Biolabs, Ipswich, USA)
- Glycoblue 15 mg/mL (Ambion)
- Lipofectamine 2000 (Invitrogen, Darmstadt)
- Lipofectamine 3000 (Invitrogen)
- Lysyl Endopeptidase (Lys-C, Wako Chemicals, Richmond, USA)
- M columns (Miltenyi Biotec)
- Mach T1 phage resistant chemically competent *E. coli* (Life Technologies)

- Nuclease-free water (Invitrogen)
- NuPage LDS Sample buffer (Invitrogen)
- Proteinase K solution (2 µg/ mL, Life Technologies)
- PVDF membrane (Immobilon-P, Millipore)
- Q5 high-fidelity DNA polymerase for cloning PCR (New England Biolabs, Ipswich, USA)
- SOC medium (super optimal broth with catabolite repression, Life Technologies)
- T4-DNA-Ligase (Roche Diagnostics, Penzberg; Germany)
- TRIzol Reagent (Invitrogen)
- Trypsin LC-MS grade 0.5 µg/µl (Promega, Madison, USA)
- X-ray films CL-XPosure (Thermo Scientific, Langenselbold, Germany)

All restriction enzymes were either obtained from Fermentas (St. Leon-Rot, Germany), New England Biolabs (Ipswich, USA) or Roche Diagnostics (Penzberg, Germany).

3.5.11. Software

- Maxquant version 1.6.0.1, 1.5.2.8, 1.5.1.2.
- R version 3.5.1 and Rstudio version 1.0.136
- Metascape
- ImageJ
- MUSCLE
- Bowtie2 (version 2.1.0)
- Tophat2 (version 2.0.10)

- Cufflinks (version 2.2.1)
- Rchie
- PhyML
- RNA decoder
- VARNA

4. METHODS

4.1. Cloning and mutagenesis

4.1.1. Templates and PCR

Gene fragments containing the coding sequence (nt 29-1007) of segment 7 of A/chicken/Rostock/45/1934 ("FPV"), A/Vietnam/1203/2004, A/swine/Netherlands/25/1980, A/BrevigMission/1/1918 ("p1918"), a mutated A/BrevigMission/1/1918 (with the following point mutations: C718T, A725C, A754G, G760A, G766A, A772G, "p1918 w/ Mal 707-779") and a mutated version of A/chicken/Rostock/45/1934 (with the following point mutations: G712A, A714G, G740A, C745T, G754A, A760G, "FPV w/ Pan 707-779") fused to attB1 and attB2 sites were ordered as synthetic double-stranded DNA fragments from Integrated DNA Technologies. The coding sequences of NS1 of Pan and Mal strain fused to attB1 and attB2 sites were codon-optimized and ordered from Gene Art AG.

The coding sequences of Pan and Mal M segments (nt 29-1007) were amplified from cDNA (obtained from Matthias Budt, RKI) and fused to attB1 and attB2 sites by PCR (polymerase chain reaction). The PCR was performed using hot-start technique (that is, the polymerase is added at the initial denaturation) in a 50 µl reaction volume. The reaction contained 1X Q5 reaction buffer, 1µl of cDNA template, 200 µM dNTP mix (New England Biolabs), 0.5 µM forward and reverse primers, 0.5 U Q5 high-fidelity DNA polymerase and nuclease-free water. The thermocycling conditions are listed in Table 3.

Table 3 Cycling conditions to amplify the M segment coding sequence for gateway cloning.

Step	Temperature	Duration	Number of cycles
Initial denaturation	98°C	5 min	1
Denaturation	98°C	30 sec	6
Annealing (Gradient)	60°C +/- 10°C	30 sec	6
Elongation	72°C	1 min	6
Denaturation	98°C	30 sec	29
Annealing	70°C	30 sec	29
Elongation	72°C	1 min	29
Final elongation	72°C	5 min	1

Samples were supplemented with *Gel loading dye* and amplicons were resolved by electrophoresis in 1 % agarose gel containing 0.4 µg/ml ethidiumbromide in *TAE buffer*. The gel was UV transilluminated and the gel band of expected size was excised with a clean scalpel. The DNA was then purified using *Invisorb Spin DNA extraction kit*.

4.1.2. Gateway cloning

Cloning was done using Gateway technology (Invitrogen) according to the manufacturer's protocol with modifications. BP-reactions were performed in 5 µl reaction volumes with 25 femtomol of PCR products (or ordered synthetic gene fragments), 75 ng of pDONR221, 1 µl of *Gateway BP clonase II enzyme mix* and nuclease free water. The reaction mixture was then incubated at room temperature for one hour. Next, 1 µg *Proteinase K* solution was added and the samples were incubated for 10 min at 37°C. Then, 2 µl were used to transform MACH1 chemical

competent cells and the cells were plated on *LB-agar* plates containing 50 µg/mL *kanamycin*. Similarly, Gateway LR reactions were performed in 5 µl reaction volumes with 25 ng of DNA from the entry clone, 50 ng of pDEST destination vector, 1 µl of *Gateway LR clonase II enzyme mix* and nuclease free water. The mixture was incubated for one hour to overnight before 1 µg of *Proteinase K solution* was added. Afterwards, 2 µl were used to transform chemically competent MACH1 bacteria and the cells were plated on *LB-Agar* plates containing 100 µg/ mL *ampicillin*.

4.1.3. Transformation and sequencing

Heat-shock transformation was done using *MACH1* chemical competent *E. coli.* DNA was pipetted to 30 µl of competent cells and incubated for 30 min at 4°C. Afterwards, cells were heated to 42°C for 1 min and then incubated for 2 min on ice. 100 µl of pre-warmed *SOC medium* was added and the cells were incubated for 30 min at 37°C on an orbital shaker. The cells were then plated onto *LB-agar* plates containing the appropriate antibiotics using a Drigalski spatula and incubated at 37°C overnight in an incubator. The following day single colonies were picked and 4 mL of *LB-medium* containing the appropriate antibiotics were inoculated with single colonies. Mini-cultures were used to inoculate 50 mL of liquid *LB-medium* containing the appropriate antibiotics.

Plasmid DNA was then purified using the commercial *Invisorb Plasmid Mini Two* kit for mini-cultures or the *Nucleobond Xtra Midi* kit for midi-cultures. Concentration and

purity was determined with spectrophotometric measurements of absorbance at 230, 260 and 280 nm wavelengths.

Pan/Mal chimeric constructs[1] were generated by replacing appropriate restriction fragments of wild type pDEST26-Flag/HA-Mseg(Pan)29-1007 and pDEST26-Flag/HA-Mseg(Mal)29-1007 respectively with the corresponding synthetic chimeric DNA inserts purchased as gBlocks from Integrated DNA Technologies. gBlocks and vectors were digested with the following enzymes: 52-739 chimera: EcoRV / NheI, 277-779 chimera: XhoI / Nhe (Mal) and BglII / NheI (Pan), 326-706 chimera: BglII / NheI (Pan) and XhoI/NheI (Mal), 707-779 chimera with BstEII / NheI. The fragments were ligated using *T4-DNA-Ligase* overnight at 16°C before transformation and extraction of plasmid DNA (see above).

4.2. RNA analysis

4.2.1. In vitro transcription

The region that is predicted to be of secondary structure in the M segment (707-825) was amplified and fused to a T3 promoter sequence with the following primers: FW 5' CCAAGCTCGAAATTAACCCTCACTAAAGG, REV: 5' GCTAGCTAGCAGTGCAAGATCCCAATG. The PCR was performed as described above (Templates and PCR), with a different cycling profile: Initial denaturation (2min, 98°C), followed by 30 cycles denaturation (30 sec, 98°C), annealing (20 sec, 70°C),

[1] Chimeric constructs were created in collaboration with Lüder Wiebusch and Barbara Vetter (Charité Berlin). Lüder Wiebusch devised the cloning strategy.

elongation (20 sec, 72°C) and a final extension step (2 min, 72°C). The PCR product was then purified and used as template for *in vitro* transcription using the *MegaScript kit* according to the manufacturer's instructions with 100 ng of template DNA. After that, template DNA was removed by adding 1 µl of TURBO DNAse mix (supplied with the kit) and incubation at 37°C for 15 min. Afterwards, the RNA was purified using phenol/chloroform according to the instructions in the kit. The RNA was then resuspended in *RNA refolding buffer* and incubated at 37°C for 30 min. To determine differences in RNA secondary structure, 500 ng of RNA was then supplemented with *gel loading dye* run on native 1 % TAE agarose gels, supplemented with ethidumbromide. To denature secondary structures, the RNA was mixed 1:5 with *RNA denaturing sample buffer* and incubated at 55°C for 15 min. Then, *RNA loading dye* was added and samples were run on a *denaturing agarose gel,* supplemented with 0.4 µg/ml ethidiumbromide. Gels were then UV transilluminated.

4.2.2. Quantitative real-time PCR[2]

Total cellular RNA was prepared using *TRIzol* reagent according to the manufacturer's instructions. RNA was quality-controlled on native agarose gels and reverse described. Prior to PCR, cDNA concentrations were adjusted to 2.5 ng/µl. Quantitative real-time PCR (qRT-PCR) was performed on a 7500 Fast Real-Time PCR System (Life Technologies) using SYBR green as dye. The gene-specific primers for M1 were: fw 5'-CTAACCGAGGTCGAAACG-3', rev 5'-CCCTTAGTCAGAGGTGAC-3'. For M2,

[2] Quantitative real-time PCR was performed by Immanuel Husic (MDC Berlin) as part of his Bachelor studies in Biotechnology.

forward primer were a 1:1 molar mixture of primers 5'-CTAACCGAGGTCGAAAC**T**CC-3' and 5'-CTAACCGAGGTCGAAAC**C**CC-3' and reverse primer: 5'-ACTCCTTCCGTAGAAGGCCC-3'. Ribosomal protein L32 (RPL32) was quantified with the forward primer: 5'-GATGCCCAACATTGGTTATGGA-3' and the reverse primer 5'-GGCACAGTAAGATTTGTTGCAC-3'. The M1/M2 mRNA levels were normalized by subtraction of RPL32 threshold cycle (C_T) values, resulting in ΔC_T values of the analyzed samples. $\Delta\Delta C_T$ values represent the fold-change between two samples in log2 space. $2^{-\Delta\Delta C_T}$ values represent the values in non-log space. Data were analyzed using the software supplied with the Fast Real-Time PCR system and plotted with in-house generated R scripts.

4.2.3. RNA-sequencing and data processing[3]

Total RNAs from infected and mock-infected A549 cells (infection conditions as described in the pAHA-SILAC, see below) were extracted using *TRIzol* reagent following the manufacturer's protocol. Truseq Stranded mRNA sequencing libraries were prepared with 500 ng total RNA according to the manufacturer's protocol (Illumina). The libraries were sequenced on a HiSeq 2000 platform (Illumina) and yielded in total 186 million 101-nt single-end reads. The sequencing reads were first subjected to adapter removal using FLEXBAR (Dodt et al., 2012) Then, the reads were mapped to the reference sequences of rRNA, tRNA, snRNA (small nuclear RNA), snoRNA (small nucleolar RNA) and miscRNAs (miscellaneous RNA) (available

[3] Library preparation, sequencing and data analysis was performed by Xi Wang, Jingyi Hou and Wei Chen. Large parts of the corresponding method section are re-worked versions of their original texts.

from Ensembl and RepeatMasker) and matching reads were excluded. The remaining reads were then mapped to the human and Pan/Mal influenza A reference genome with known viral annotation and the RefSeq/Ensembl human gene structure. Gene expression levels (reads per kilobase per million reads, RPKM) were estimated by Cufflinks (v2.2.1) (Trapnell et al., 2010). Splice junction reads for various M transcripts were counted with customized Perl scripts. Splice isoforms were accepted that had >500 read counts in both replicates.

4.2.4. Computational analysis of RNA secondary structure[4]

M segment sequences were obtained from the NIAID Influenza Research Database (IRD) (Zhang et al., 2017) through the web site at http://www.fludb.org. We used the following settings for human-adapted strains: date range >= 2009, sub-type H3N2, only complete genomes, host human; exclude laboratory strains and duplicate sequences, geographic grouping: South America, Europe and Asia; for avian-adapted strains: date range >= 2009, only complete genomes, host: avian; exclude duplicate sequences and geographic grouping: Europe and Asia. These two sets of sequences were merged with the respective reference sequences (i.e. Pan and Mal). Alignments were performed using the program Muscle (Edgar, 2004) and comprised 403 sequences for the human-adapted and 199 sequences for the avian-adapted strains. The program PhyML (Guindon et al., 2009) was used in conjunction with the HKY evolutionary model generate evolutionary trees.

The two input alignments (including the combined annotation of the known protein-coding M1 and M2 regions) and the corresponding evolutionary trees were then used as input to RNA-Decoder (Pedersen et al., 2004). The predictions by RNA-Decoder gave the posterior base-pairing probabilities for each base-pair of the predicted RNA structure. Predicted base-pairs with a probability below 25 % were omitted from the RNA structure visualization that were performed with R-chie (Lai and Meyer, 2014). Finally, the predicted RNA structure element (nt 733-766) was plotted with the sequence of the Mal strain using VARNA tool (Darty et al., 2009).

[4] Alignment, evolutionary trees and the prediction of RNA secondary structure were performed by Irmtraud Meyer. The corresponding Methods section is a re-worked version of her original text.

4.3. Cell biology

4.3.1. Cells and viruses

A549 and HEK293T cells (human) were grown in *DMEM culturing medium 1*. MDCK type II cells were grown in *MEM culturing medium*. DF-1 cells (chicken) were grown in *DMEM culturing medium 2*. All cells were maintained at 37 °C and 5 % CO_2. Stocks of the avian influenza virus A/Mallard/439/2004 (H3N2) (Mal) were grown in the allantoic cavities of 10-day-old embryonated chicken eggs for 2 days at 37°C. A/Panama/2007/1999 (H3N2) (Pan), Pan + Mal M reassortant and Pan-Av mutant virus were grown in MDCK type II cells. Virus stocks were titrated on MDCK type II cells by measuring plaque forming units (PFU) according to standard protocols or fluorescence forming units (FFU, see below).

4.3.2. Titration of IAV viruses using fluorescence forming units

For titration, MDCK type II cells were infected with different dilutions of virus stocks for 5 h. Then cells were harvested by trypsinization, fixed and permeabilized by incubation in 75 % ethanol for at least 12 h at 4°C and stained with specific antibody against NP antigen. An Alexa Fluor 488-conjugated goat anti-mouse IgG antibody (Invitrogen) was used as secondary reagent. Cells were analyzed on a FACSCanto II flow cytometer (BD Biosciences) using FACSDiva software package (BD Biosciences).

To gate NP-positive cells, dot plots were created displaying on a linear scale the forward light scatter (FSC) and sideward light scatter (SSC) of measured particles (Figure 7A,C). An FSC threshold was set for exclusion of cell debris. A region R1 was set that excludes cell doublets and aggregates from further analysis. The R1 population was analyzed on a separate dot plot for FITC fluorescence of infected cells (Figure 7B,D). Non-infected cells were used to discriminate FITC-positive, NP-expressing cells from background fluorescence. A region R2 was set to gate only FITC-positive cells and calculate their proportion of the parental population R1. This percentage was used to calculate the fluorescence forming units (FFU) in the virus inoculum. For the calculation, only samples were used that had a percentage of NP-positive cells between 1 % and 15 %.

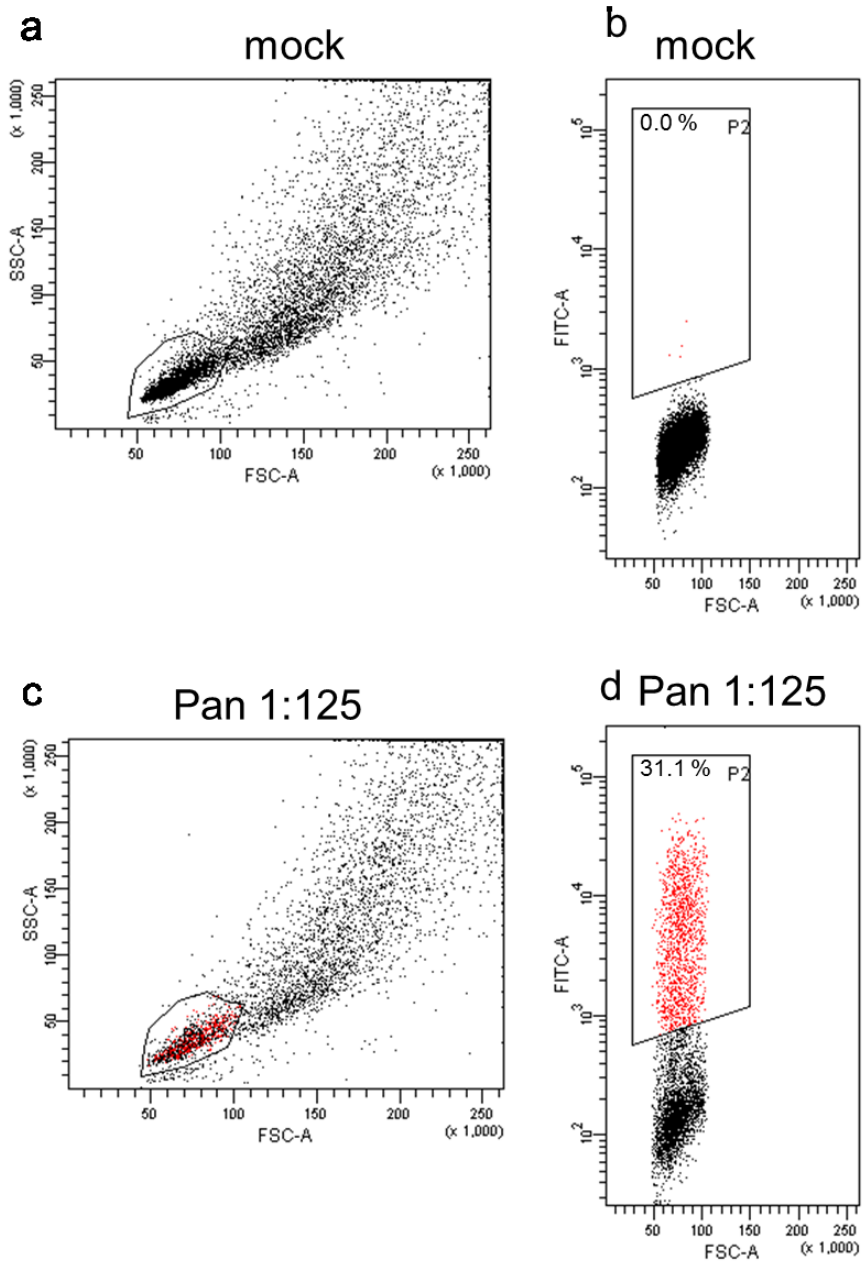


Figure 7 Gating strategy to determine the amount of infectious units in the inoculum.

(a,c) *An FSC-SSC dot plot was created to remove doublets and aggregates from analysis.* ***(b,d)*** *This subset was assessed in FITC-channel to discriminate non-infected (black) from infected cells (red). Representative dot plots are shown.*

4.3.3. Reverse genetics[5]

A cDNA copy of the M segment of A/Mallard/439/2004 was amplified by RT-PCR followed by insertion into pHW2000 resulting in pHW2000-Mal M. Also, the pHW2000-Pan M plasmid was mutated by introduction of eight point mutations into the M1 reading frame within pHW2000-M (A712G, G714A, A740G, T745C, A754G, G760A, G766A, A772G, pHW2000-Pan-AV). Likewise, the pHW2000-Mal M plasmid was mutated by introduction of eight point mutations into the M1 reading frame within pHW2000-M (G712A, A714G, G740A, C745T, G754A, A760G, A766G, G772A, pHW2000-Mal-Hu). Both pHW2000 constructs were mutated using the *Gibson Assembly cloning kit* according to the manufacturer's instructions.

The pDEST26-Flag/HA-Pan w/ Mal 707-779 and the wild type pHW2000-Pan M served as template for PCR. All constructs were confirmed by cycle sequencing.

Recombinant influenza A viruses were derived from the A/Panama/2007/99 backbone and generated using an eight plasmid system for this strain based on pHW2000 plasmids. In brief, human HEK293T were transfected using *Lipofectamine 2000* according to the manufacturer's protocol. Then, supernatant was passaged on MDCK type II cells as described previously (Matthaei et al., 2013). For the rescue of the Pan-Mal M reassortant virus, we used pHW2000-Mal M, whereas the Pan-AV virus was generated with pHW2000-Pan-AV together with seven plasmids encoding the other segments of human A/Panama/2007/99 virus.

[5] The reconstitution of Pan-AV mutant virus and of the Pan-Mal-M reassortant was performed by Gudrun Heins and Thorsten Wolff, Robert-Koch-Institute. Their original texts have been modified for this purpose.

4.3.4. Virus infections

All infectious work was carried out under BSL2 (biosafety level 2) conditions. For infection, eukaryotic cells were seeded one day prior to the experiment so that they were confluent on the day of infection. Then cells were washed in *PBS*$^{+/+}$. Then, virus stocks were diluted in *PBS*$^{+/+}$ to the appropriate MOI (multiplicity of infection) and used to infect cells for 45 min at room temperature, unless stated otherwise. Mock-infection was performed with *PBS*$^{+/+}$ only. The volume of infection was chosen as the minimal volume that allowed complete coverage of the cells. Then cells were washed in *PBS* and appropriate infection medium was added. Cells were harvested by scraping before they were washed once with *PBS*. Then they were subjected lysis and immunoblotting or qRT-PCR.

4.3.5. Immunofluorescence microscopy

A549 were grown on glass coverslips and infected with the indicated viruses at an MOI of 1 (FFU/cell). At the indicated time points post infection, cells were washed twice with PBS and then fixed for 15 min in *Fixation solution.* The cells were washed again twice and permeabilized in *Permeabilization solution* before being washed again twice and stained in 50 µl *Staining solution 1.* After washing the cells were stained with secondary antibody in *Staining solution 2.* After washing in ddH_2O the coverslips were put on slides using *Mowiol solution*. Images were acquired by an Eclipse A1 laser-scanning microscope using the NIS-Elements software package (Nikon). At least 150 cells were counted per condition to quantify the subcellular distribution of NP using ImageJ software (Schneider et al., 2012).

4.3.6. Transfections and lysis

A549 cells were seeded on 6-well plates and transfected with 2.5 µg of expression constructs using *Lipofectamine 3000* (Thermo Fisher) reagent according to the manufacturer's instructions. Cells were harvested 24 hours post transfections by trypsinization and subjected to lysis and immunoblotting or qRT-PCR. For the RdRp reporter system, HEK293T cells were transfected in 6-well dishes with pHW2000 constructs encoding for PA, PB1, PB2 (each 0.7 µg DNA), NP (1.4 µg DNA) and wild type and chimeric M (0.35 µg DNA) segments. Transfections were done using *PEI solution* (15 µl) for each transfection setup. Cells were harvested 24 hours post transfection and subjected to lysis and immunoblotting.

For AP-MS experiments, fully SILAC labeled HEK293T cells were seeded on 15 cm plates one day prior to infection. On the day of transfection plasmid DNA (15 µg) and *PEI solution* (45 µl) were used to transfect cells. After incubation for 24 hours cells were subjected to affinity purifications followed by mass-spectrometry or lysis and immunoblotting.

If not mentioned otherwise, cells were lysed by resuspending a cell pellet with an appropriate volume of *lysis buffer*. The lysate was incubated for 30-45 min at 4 °C while rotating head-over-tails. Afterwards, the lysate was cleared by centrifugation for 20 min.

4.4. Protein biochemistry

4.4.1. Affinity purification

For anti-c-myc or anti-HA affinity purification of myc-NS1 or Flag/HA-NS1 of Pan and Mal strain, a *µMACS c-myc / HA isolation kit* was employed according to the manufacturer's instructions with modifications using *M columns*. To avoid slow running through the columns, all buffers were degassed before use. Cleared lysates were incubated with magnetic beads for 45 min. The lysate bead mixture was then applied to the column. The first washing step was carried out using *lysis buffer*. As detergents can be detrimental to mass-spec analysis a *lysis buffer* without detergents was used for the following washing step. The final washing step was carried out in 20 mM Tris-HCl, pH 7 (supplied with the kit).

Samples destined for mass-spectrometry were eluted in a total volume of 0.2 ml 8 M guanidine hydrochloride at 95°C. Proteins were precipitated from the eluates by adding 1.8 ml LiChrosolv ethanol and 1 µl *GlycoBlue* and incubation at 4°C overnight. Samples were centrifuged for 1h at 4°C and ethanol was decanted. The protein pellet was air-dried and subjected to in-solution digestion. Samples destined for immunoblotting were eluted with 100 µl pre-heated (95°C) hot Elution buffer (supplied with the kit) and subjected to SDS-PAGE and Immunoblotting.

4.4.2. SDS-PAGE and immunoblotting

Samples were supplemented with *NuPage LDS Sample buffer*, 50 mM DTT and heated for 10 min at 70°C. Afterwards, proteins were resolved at constant voltage

(100 V) on *4 %–12 % Bis-Tris gradient gels* before being blotted onto *PVDF membranes* using a wet blotting system (Invitrogen) for 2 hours at constant current (250 mA) in *Transfer buffer*. The membrane was blocked in *Blocking buffer*, followed by incubation with primary antibodies diluted in *Blocking buffer* overnight. After three 5 minute-long washes in *TBST* the membrane was incubated with appropriate secondary antibodies diluted in *Blocking buffer* at room temperature for 1 hour. Afterwards the membrane was incubated with substrates for chemiluminescence (*Novex ECL HRP chemiluminescent substrate reagent kit*). The signal was detected either with a ChemiDoc MP Imaging System (Bio-Rad, München, Germany) or with X-ray films.

4.5. pAHA-SILAC

4.5.1. Sample preparation

A549 cells were fully labeled in *SILAC DMEM* containing either light, medium or heavy lysine and arginine. Cells were cultured in *SILAC DMEM* for at least 6 passages. 10 cm dishes of confluent light labeled cells were mock-infected, while heavy and medium labeled cells were infected with either Pan or Mal strain at an MOI of 3 (PFU). Virus was allowed to attach to the cells for 45 min on ice. Cells were washed with pre-warmed *PBS* before 5 mL *Infection SILAC DMEM* was added. Thirty minutes prior to pulse-labeling cells were washed with pre-warmed $PBS^{+/+}$ and incubated in 5 mL *Infection SILAC DMEM w/o methionine*. During the pulse interval cells were incubated in 5 mL *Infection SILAC DMEM w/o methionine* additionally containing 100 µM L-Azidohomoalanine. Cells were washed in *PBS*, scraped from the dish and frozen at -80°C until further sample processing. Lysis and enrichment for newly synthesized proteins was done using *Click-iT Protein Enrichment Kit*. In detail, the Urea lysis buffer supplied with the kit was supplemented with 1 tablet of *protease inhibitor* per 5 mL. 283 µl of urea lysis buffer was used per label and cells were incubated for 20 min on ice before being sonicated twice for 10 seconds with 5 minutes pauses in between the sonication pulses. Cell debris was removed by centrifugation at maximum speed in a benchtop centrifuge before SILAC label were mixed. 10 % of sample were directly subjected to Wessel-Fluegge precipitation (Wessel and Flugge, 1984)) and served as the input, 90 % were used for enrichment of newly synthesized proteins as previously described (Eichelbaum and Krijgsveld, 2014a) with modifications (see below).

4.5.2. Enrichment of newly synthesized proteins

In brief, for the enrichment of newly synthesized protein 200 µl of alkyne agarose resin per enrichment reaction was prepared by first washing it with ddH_2O. Then 800 µl of cell lysate (from the mixed labels) was added to the beads together with 1 mL of 2X catalyst solution (containing 835 µl ddH_2O, 125 µl reaction additive 1, 20 µl of 100 mM copper-II-sulfate and 20 µl of reaction additive 2). The enrichment reactions were then rotated head-over-tails for 18 hours at room temperature. Enriched proteins were washed twice in 1.8 mL ddH_2O before being reduced in 1 mL SDS wash buffer additionally containing 10 µl of 1 M DTT. The resin was heated to 70°C on a heating block for 15 min and then cooled to room temperature for 15 min. Proteins were alkylated for 30 min in the dark in SDS wash buffer containing 40 mM iodoacetamide. Beads were then washed sequentially (each 5x) in SDS wash buffer, 8 M urea in 0.1 M Tris/HCl (pH 8.0), 80 % acetonitrile in 0.1 M Tris/HCl pH 8.0 and 5 % acetonitrile in 50 mM ammonium bicarbonate. Proteins were then digested in 5 % acetonitrile / 50 mM ammonium bicarbonate overnight using 1 µg *trypsin* (Promega). Peptides were then acidified using *TFA*, desalted and either directly measured on a nano LC-MS/MS set-up (see below: 4.5.4. Mass spectrometry) or subjected before to isoelectric focusing using an OFFGEL-fractionator 3100 (Agilent, Santa Clara, USA).

4.5.3. Input preparation

Input samples were reduced by adding DTT to a final concentration of 0.1 M and incubation for 5 min at 95 °C. Sulfhydryl groups were alkylated by adding

iodoacetamide to a final concentration of 0.25 M and incubation for 20 min in the dark at room temperature. Proteins were precipitated according to Wessel and Fluegge (Wessel and Flugge, 1984), resuspended in 6 M urea / 2 M thiourea and digested into peptides with C-terminal lysine or arginine using Lys-C (3 h) and trypsin (overnight, diluted 4× with 50 mm ABC). Enzyme activity was quenched by acidification of the samples by adding 10 µl *TFA*. The peptides were desalted with in-house prepared C18 Stage Tips (Rappsilber et al., 2003) prior to nano LC-MS/MS analysis.

4.5.4. Mass spectrometry

Peptides from input and AHA-enriched samples were separated on a monolithic silica capillary column (MonoCap C18 High Resolution 2000, GL Sciences), 0.1 mm internal diameter × 2,000 mm length, at a flow rate of 300 nL/min with a 5 to 45 % acetonitrile gradient on an EASY-nLC II system (Thermo Fisher Scientific) with 480 or 240 min gradient or on a EASY-nLC HPLC (Thermo Fisher) system by 2 or 4 h gradients with a 250 nl/min flow rate on a 15 cm column with an inner diameter of 75 µm packed in house with ReproSil-Pur C18-AQ material (Dr. Maisch, GmbH). Peptides were ionized using an ESI source on a Q-Exactive, Q-Exactive Plus or a LTQ (linear trap quadrupole) Orbitrap Velos MS (all Thermo Fisher) in data dependent mode. Q-Exactive and Q-Exactive Plus mass spectrometers were operated in the data dependent mode with a full scan in the Orbitrap followed by top 10 MS/MS scans using higher-energy collision dissociation. The full scans were performed with in a m/z range of 300 - 1,700, a resolution of 70,000, a target value of 3×10^6 ions and a maximum

injection time of 20 ms. The MS/MS scans were performed with a 17,500 resolution, a 1×10^5 target value and a 60 ms maximum injection time. The LTQ Orbitrap Velos instrument was operated in data dependent CID top 20 mode. Full scans were performed in m/z range 300-1,700 with a resolution of 60,000 and a target value of 10^6. MS/MS scans were performed with an isolation window of 2 m/z and a target value of 3,000.

4.6. pSILAC/ AP-MS

4.6.1. pSILAC sample preparation

Cells were adapted to *SILAC DMEM light* medium one day before the experiment, seeded in 6-wells and infected as described above (Virus infections). Prior to the pulse period, cells were maintained in $PBS^{+/+}$ for 30 min to starve them of light amino acids. Then cells were pulse labeled with *pulsed SILAC Infection DMEM medium or heavy* for the indicated time intervals, harvested and combined. Lysis was carried out in *lysis buffer* for 1 h on a rotating wheel with subsequent centrifugation. Proteins were then subjected to Wessel-Fluegge precipitation (Wessel and Flugge, 1984) and in solution digestion.

4.6.2. In-solution digestion

Protein pellets obtained from Wessel-Fluegge (Wessel and Flugge, 1984) or ethanol precipitation were first resolved in *denaturation buffer*. DTT was added to a final concentration of 10 mM and incubation for 30 min. Reduced thiol groups were then alkylated by adding 1:10 volume of *alkylation buffer* and incubation for 30 min in the dark at room temperature. Proteins were digested into peptides with C-terminal lysine or arginine using Lys-C (3 h, room temperature, 1 µg per 50 µg protein) followed by trypsin (overnight,, diluted 4X with *ABC buffer*). Enzyme activity was quenched by acidification of the samples with *TFA*. The peptides were desalted with C18 Stage Tips according to standard protocols (Rappsilber et al., 2003) prior to nano LC-MS/MS analysis.

4.6.3. Mass spectrometry

Peptides from pSILAC samples were separated by 4 h gradients and ionized with ESI source and analyzed on Q-Exactive HF-X instrument (Thermo Fisher) in data dependent mode. The full scans were performed with a resolution of 60,000, a target value of 3×10^6 ions and a maximum injection time of 10 ms. The MS/MS scans were performed with a 15,000 resolution, a 1×10^5 target value and a 22 ms maximum injection time.

Peptides from AP-MS samples were separated by 2 h gradients and ionized with ESI source and analyzed on Q-Exactive Plus instrument (Thermo Fisher) in data dependent mode. The full scans were performed with a resolution of 70,000, a target value of 1×10^6 ions and a maximum injection time of 120 ms. The MS/MS scans were performed with a 17,500 resolution, a 1×10^5 target value and a 60 ms maximum injection time.

4.7. Proteomic data analysis

4.7.1. MaxQuant analysis

Raw files for pAHA-SILAC were analysed with MaxQuant (Cox and Mann, 2008) software version 1.6.0.1 Default settings were kept except that 'requantify' option was turned on. Label-free quantification via iBAQ calculation was enabled. Lys4/Arg6 and Lys8/Arg10 were set as labels and oxidation of methionines, N-terminal acetylation and deamidation of asparagine and glutamine residues were defined as variable modifications. The in silico digests of the human Uniprot database (downloaded January 2018), the protein sequences of twelve Pan and Mal Influenza virus proteins and a database containing common contaminants were done with Trypsin/P. The false discovery rate was set to 1 % at both the peptide and protein level and was assessed by in parallel searching a database containing the reverted sequences from the Uniprot database. The resulting text files were filtered to exclude reverse database hits, potential contaminants and proteins only identified by site (that is protein identifications that are only explained by a modified peptide). In all cases, SILAC ratios from the label-swap experiments were inverted. Plotting and statistics were done using R and figures were compiled in Illustrator (Adobe). Raw files for pSILAC were analysed as described above, except that MaxQuant software version 1.5.2.8 was used and requantify option was set to off for the dataset comparing Pan and Pan-AV strain. Raw files for pSILAC comparing the Mal strain in DF-1 and A549 cells were searched with a combined uniprot database of *gallus gallus* and *homo sapiens*. 'Requantify' option was set to on, to allow quantification of M2 SILAC pairs. Raw-files

from the AP-MS data were searched as the pAHA-SILAC dataset with MaxQuant software version 1.5.1.2 and enabled re-quantify option and disabled iBAQ option.

4.7.2. Data processing

In general, two MaxQuant output files were used: the “proteinGroups.txt” file and the “evidence.txt” file. The latter contains all information on the identified peptides, including peptide sequence, protein ID, modification status, search score, *m*/*z*, mass-error, charge, peptide ratio quantifications etc. The former contains information on the identified protein groups, including number of identified peptides, ratio quantifications, iBAQ values, protein intensities (sum of all peptide intensities), etc.

For the pAHA-SILAC data, the iBAQ values from infected samples were extracted from proteinGroups.txt. iBAQ values were first normalized by scaling to the iBAQ protein median across all mock infected samples. This step is necessary as samples from different time periods were measured with different gradient lengths. This step also assumes that there are no differences in overall protein synthesis between different mock-infected samples. The iBAQ values were averaged for the corresponding label-swap replicates and proteins were categorized as host or viral. For estimating the newly synthesized protein mass, intensity values of H and M SILAC channels (infected with either Pan or Mal strain) were divided by the summed up intensities of the SILAC L channel (mock infected). Data was then averaged for label swap replicates and summed up for viral and host proteins independently. Finally, data was normalized to the 0-4h time period (i.e. the 0-4 h time period was set to 1).

SILAC ratios of host proteins were first transformed into log2 space. The median SILAC H/L and SILAC M/L ratios from the input samples were used to estimate the mixing ratio of the input and the H/L and M/L ratios after the enrichment were adjusted correspondingly. Figure 8 exemplarily depicts this normalization strategy.

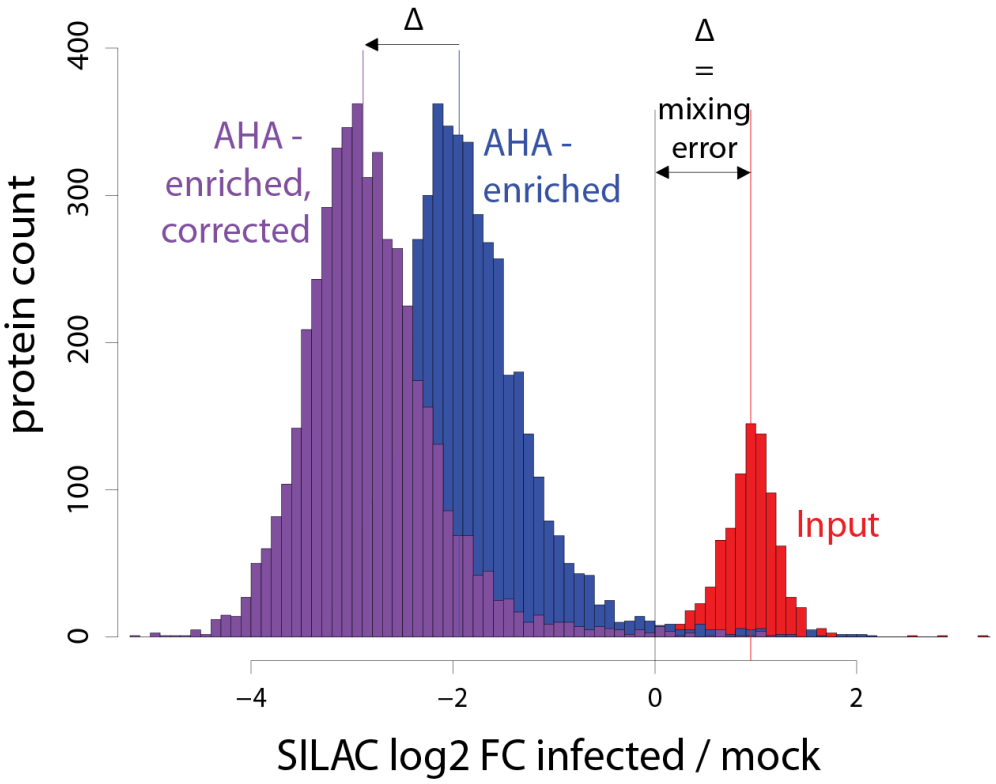


Figure 8 Normalization of pAHA-data.

SILAC ratio of protein measured in the input samples (red) serve to determine the mixing errors of pAHA-SILAC experiment. AHA-enriched ratios (blue) were adjusted by subtracting the median of the input ratios from AHA-enriched ratios, yielding corrected AHA-enriched ratios (purple).

SILAC H/M ratios that relate to the Pan/Mal (or Mal/Pan) infection treatment were normalized to 0. Afterwards, the replicate measurements were averaged. Proteins that were quantified in only one replicate were excluded. Then, the top 2 % of proteins (highest log2 fold-change) of either Pan/Mock or Mal/Mock condition were selected

and multi-set GO (gene ontology)-enrichment was performed using Metascape tool (http://metascape.org) (Tripathi et al., 2015).

Protein level data was matched to RNA level data based on the HGNC official gene symbol. Protein synthesis efficiencies were calculated by subtracting log10(RPKM) from log10(iBAQ) values.

For quantification of pAHA-SILAC viral protein expression kinetics, we extracted all quantifications of peptide level evidences for each individual replicate from evidence.txt file (for the H/M and M/H ratios that relate to the Pan/Mal infection condition). Median log2 ratios were then normalized to 0 for individual replicates. Comparative viral protein expression kinetics were based on Pan/Mal shared peptides. For each time period, replicate and viral protein the Pan/Mal SILAC protein ratio was calculated as the median of all SILAC peptide level ratios. Replicate SILAC protein ratios were averaged.

For pSILAC, ratios for viral proteins were extracted from proteinGroups.txt. Non-normalized and log2-transformed ratios were used in both experiments (Pan versus Pan-AV) and Mal (A549) versus Mal (DF-1). Quantifications of viral proteins with >75 % ratio variability were removed.

For AP-MS data, normalized SILAC ratios were extracted from proteinGroups.txt and log2 transformed.

4.8. Retrieval of M segment consensus sequences

To retrieve M segment consensus sequences for human and avian-adapted strains, we used the NIAID Influenza Research Database (IRD) (Zhang et al., 2017) through the web site at http://www.fludb.org (downloaded 10th of July 2018). The following options were applied to retrieve avian-adapted consensus sequences: date-range: 1999-2011, complete genome, minimum length M segment sequence 1027, exclude duplicated and laboratory strains. The following options were used to retrieve the human consensus: date-range: 1999-2011, complete genome, minimum length M segment sequence 1002, exclude duplicated and laboratory strains. The alignment was performed with the Alignment tool on the FluDB website. This tool allows only for 1,000 sequences to be aligned, which is the reason for applying the above mentioned segment length constraints.

5. RESULTS[6]

5.1. Experimental approach

To compare the kinetics of protein synthesis upon infection with both strains we performed proteome-wide comparative pulse-labeling experiments by combining labeling with AHA (see 1.2.2. Assessment of proteome dynamics) and SILAC (Figure 9a): This strategy allows to quantitatively monitor the kinetics of protein synthesis with high temporal resolution (Eichelbaum and Krijgsveld, 2014a). In the first step, human lung adenocarcinoma cells (A549) are fully labeled using SILAC. Second, SILAC-labeled cell populations were infected with either Pan or Mal strain or left uninfected. Third, all cells were pulse labeled with AHA for four hours during four time intervals (0-4, 4-8, 8-12 and 12-16 h). We used relatively short pulse intervals because AHA has been shown to have adverse effect on protein homeostasis and overall protein synthesis when given to cells for prolonged periods (Bagert et al., 2014).

The three cell populations for every time interval were then lysed, combined and AHA-containing proteins were enriched from the mixed lysate using click chemistry (Figure 9b). After on-bead digestion, peptide samples were analyzed by high resolution shotgun proteomics. We quantified proteins using two readouts: (i) SILAC-based relative quantification to assess differences in *de novo* protein synthesis of individual

[6] The results section contains data that was compiled and submitted as a manuscript under the title "The dynamic proteome of Influenza A virus infection identifies M segment RNA splicing as host range determinant" for review. At the time of writing this chapter, the manuscript was uploaded to biorxiv.org and is in revision at *Nature Communications* (available under the following link:https://www.biorxiv.org/content/10.1101/438176v1). All results that were obtained in collaboration with other researchers are mentioned in the figure legends. In addition, contributions made by others to individual results of the thesis are mentioned in the statement of contribution (page vii).

protein comparing infection conditions and (ii) intensity-based absolute quantification (iBAQ) to quantify absolute amounts of newly synthesized proteins (Schwanhausser et al., 2011). Thus, we recorded kinetic profiles for relative and absolute differences in *de novo* protein synthesis across the course of infection. In total, we identified 7,189 host and 10 viral proteins and quantified 6,019 proteins in at least two biological replicates (Supplementary Table 1). Label-swap replicates were overall good reproducible, with Spearman's correlation coefficients ranging from 0.44 to 0.8 (Supplementary Figure 1 and 2).

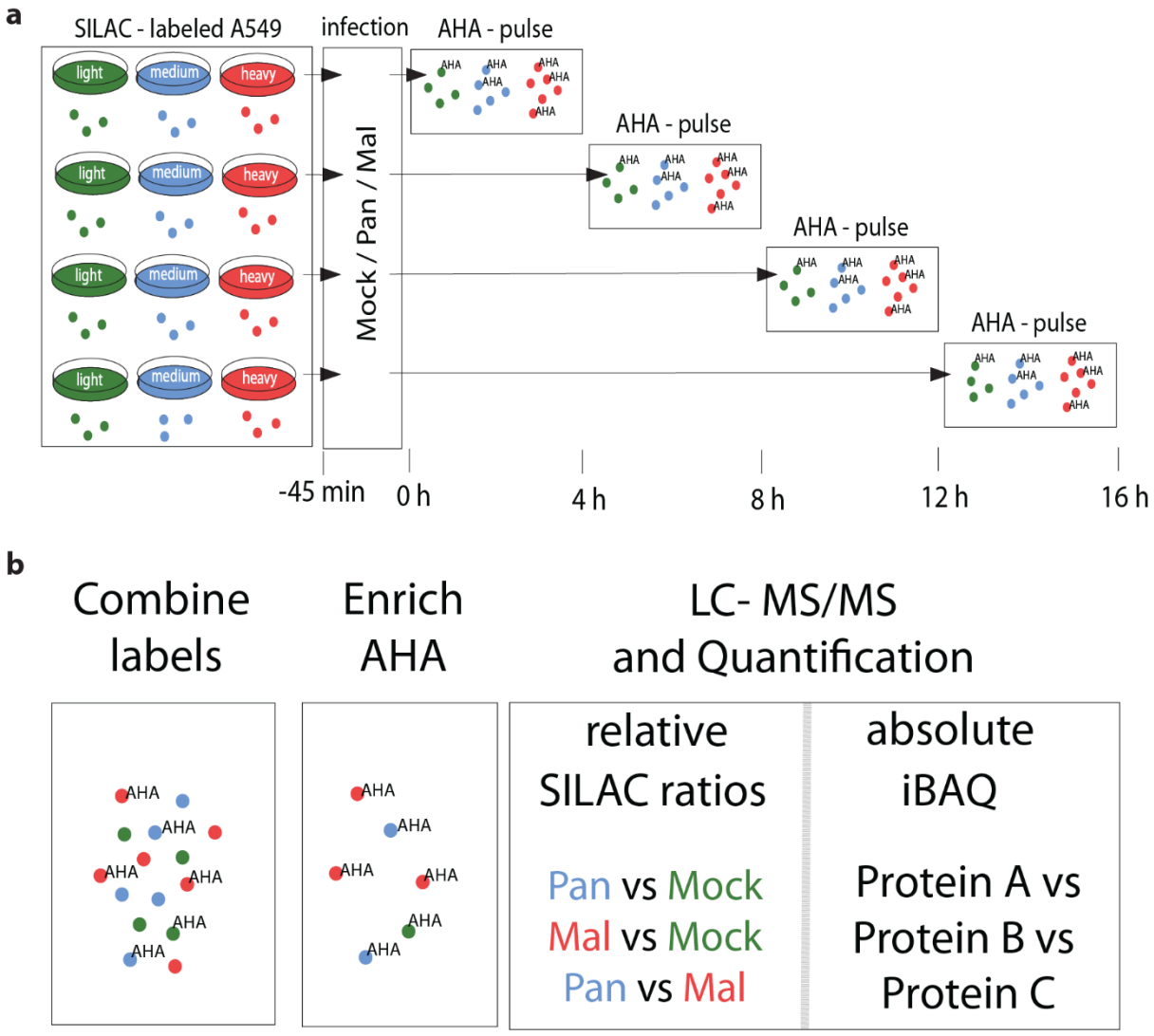


Figure 9 A strategy to quantify protein de novo synthesis proteome-wide upon IAV infection.

(a) *SILAC L/M/H - labeled A549 cells were infected with the human seasonal H3N2 IAV isolate Panama (Pan) or the avian H3N2 isolate Mallard (Mal) or left uninfected. The methionine analogous AHA was given to the methionine-depleted medium in different 4 h intervals. Importantly, the set-up was designed in label-swap duplicates. While the SILAC-light cells were mock infected in both replicates, Mal was used to infect SILAC heavy cells in replicate 1 and SILAC medium cells in replicate 2. Likewise, SILAC medium cells were Pan infected in replicate 1, while SILAC heavy cells were Pan infected in replicate 2.* ***(b)*** *After lysis and enrichment for proteins that incorporated AHA, samples were subjected to shotgun-proteomics. Absolute and relative protein synthesis profiles were quantified for host and viral proteins.*

5.2. The dynamic host proteome

5.2.1. pAHA-SILAC captures host shutoff

Many, perhaps most, viruses induce a global reduction in the synthesis of host proteins. IAV is a prime example for this phenomenon that is called “host shutoff”. The host shutoff is thought to be caused by several viral effectors, including NS1, PA-X, and PA (Rivas et al., 2016). To assess this shutoff in our proteomic data, we looked at the absolute levels of newly synthesized proteins by analyzing iBAQ values. An iBAQ value is calculated as the sum of all peptide intensities for a protein divided by the number of tryptic peptides (Schwanhausser et al., 2011). These values correlate with the protein copy number (Schwanhausser et al., 2011).

In our data we observed that, as expected, viral proteins were potently induced while the production of host proteins decreased over time (Figure 10a,b). The difference between both subsets of proteins reached several orders of magnitude (~200 fold) and peaked during the 8-12 h pulse period. Interestingly, the magnitude of host shutoff was comparable between both isolates, suggesting that the principle mechanisms that drive shutoff are in effect during permissive and non-permissive infection. In addition, the total protein synthesis output produced by the virus and the cell dropped to ~24 % (Pan) or ~30 % (Mal) at later stages of infection. Of this, ~20-40 % was of viral origin (Figure 10c,d). Thus, IAV infection is accompanied by host shutoff and global reduction of cellular protein synthesis output with no major differences comparing

permissive and non-permissive infection. This suggests that the principal mechanisms driving host shutoff are conserved for both IAV isolates.

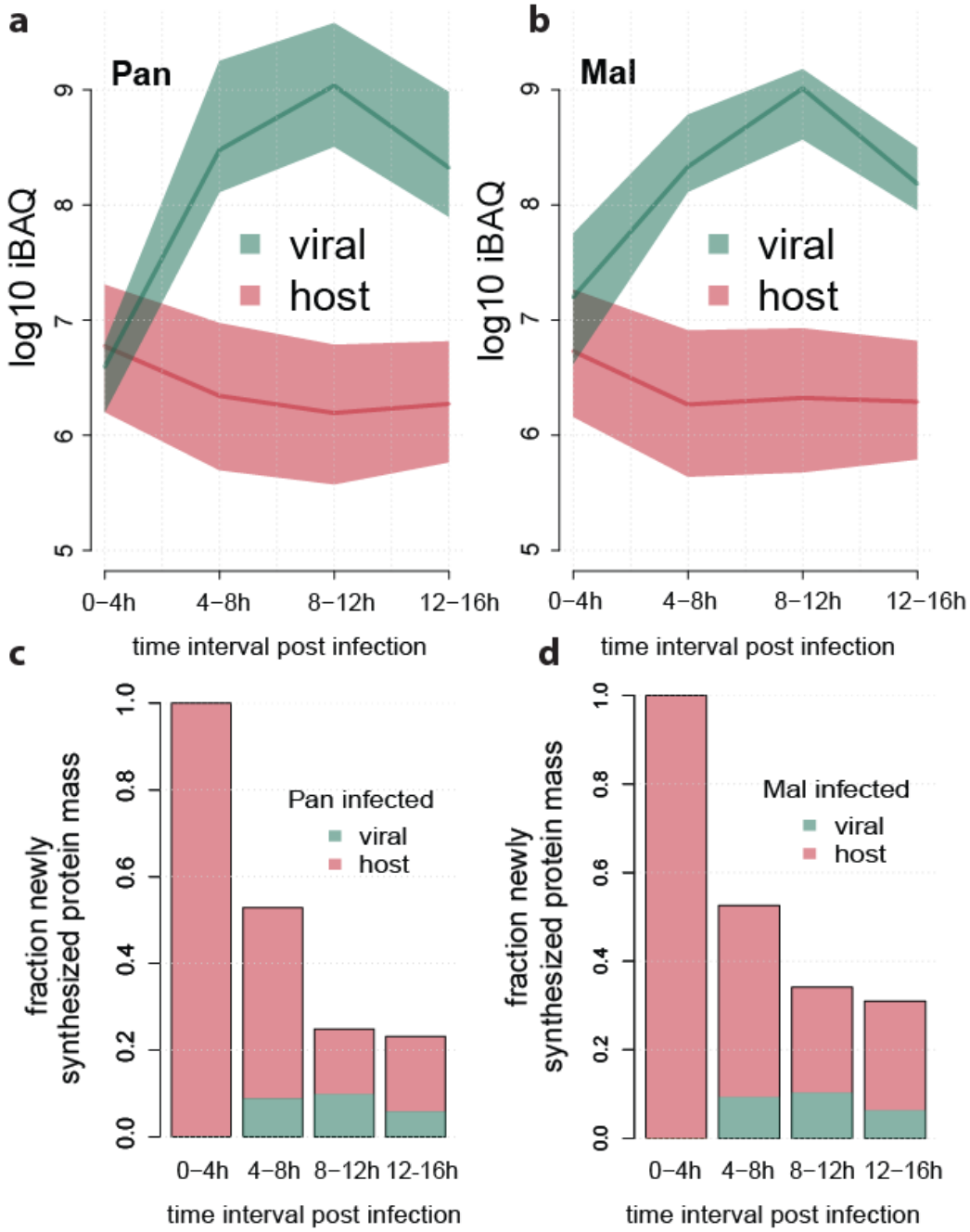


Figure 10 Quantification of host shutoff.

*iBAQ based quantification of protein synthesis levels for host and viral proteins in cell infected with either Pan (**a**) or Mal (**b**) virus, as indicated. Median, 25th and 75th percentiles of the respective populations are given. Quantification of the total newly synthesized protein mass for host and viral proteins with either Pan (**c**) or Mal (**d**) infection as indicated. Data was normalized to the 0-4 h time period. All data depict the average of two replicates.*

5.2.2. Subsets of host proteins escape IAV induced host shutoff

Our quantitative approach with SILAC allowed us to relatively compare the synthesis of individual proteins between different infection conditions. Hence, we investigated the profiles of individual host proteins across the progress of infection. For this, we directly looked at SILAC ratios comparing infected and non-infected cells (Figure 11). We observed that the synthesis of most proteins decreased strongly over time. This is expected because of the IAV-induced host shutoff. In contrast, the synthesis of some proteins was less affected and some even increased in their synthesis during later stages (Figure 11a). We aimed to systematically assess which proteins were least subject of host shutoff and selected the proteins with the highest log2 fold-changes. We did this for all time intervals, performed multi-list GO enrichment (Figure 11b) and then compiled the data into a heatmap. The heatmap presents an overview of biological processes that were least subject to shutoff as the infection progresses. For example, many interferon stimulated genes (ISGs) were relatively strongly produced at intermediate to late stages post infection (e.g. oligoadenylate synthetases, MX1, IFIT proteins (interferon induced proteins with tetraticopeptide repeats)). In addition, we observed that ribosomal proteins (GO: “peptide chain elongation”) were induced at intermediate to late time intervals. Interestingly, proteins related to mitochondria (mitochondrial ribosomal, respiratory chain proteins) and steroid metabolism escaped the shutoff at early to intermediate stages post infection. The response at the protein synthesis level was overall similar for both strains. To assess potential more subtle differences between strains, we selected a set of GO terms and compared host proteins related to the respective GO term directly between

both strains (Figure 11c). We found that proteins relating to “steroid metabolic process” and “mitochondrial gene expression” where overall produced similarly between both strains. In contrast, proteins relating to the “Type I interferon response” were first stronger produced by the avian virus. Later on, the human virus elicited higher synthesis of interferon-related proteins. This is in consistence with our previously published data obtained with interferon-β ELISAs comparing Mal and Pan infection in A549 cells (Sadewasser et al., 2017). Collectively, these data highlight subsets of proteins least subject to host shutoff. Among the few differences that we observed between both strains is the induction of IFN-sensitive genes. Further, this suggests that it is not an increased response to type I IFNs that limits the ability of the avian strain to multiply in human cells.

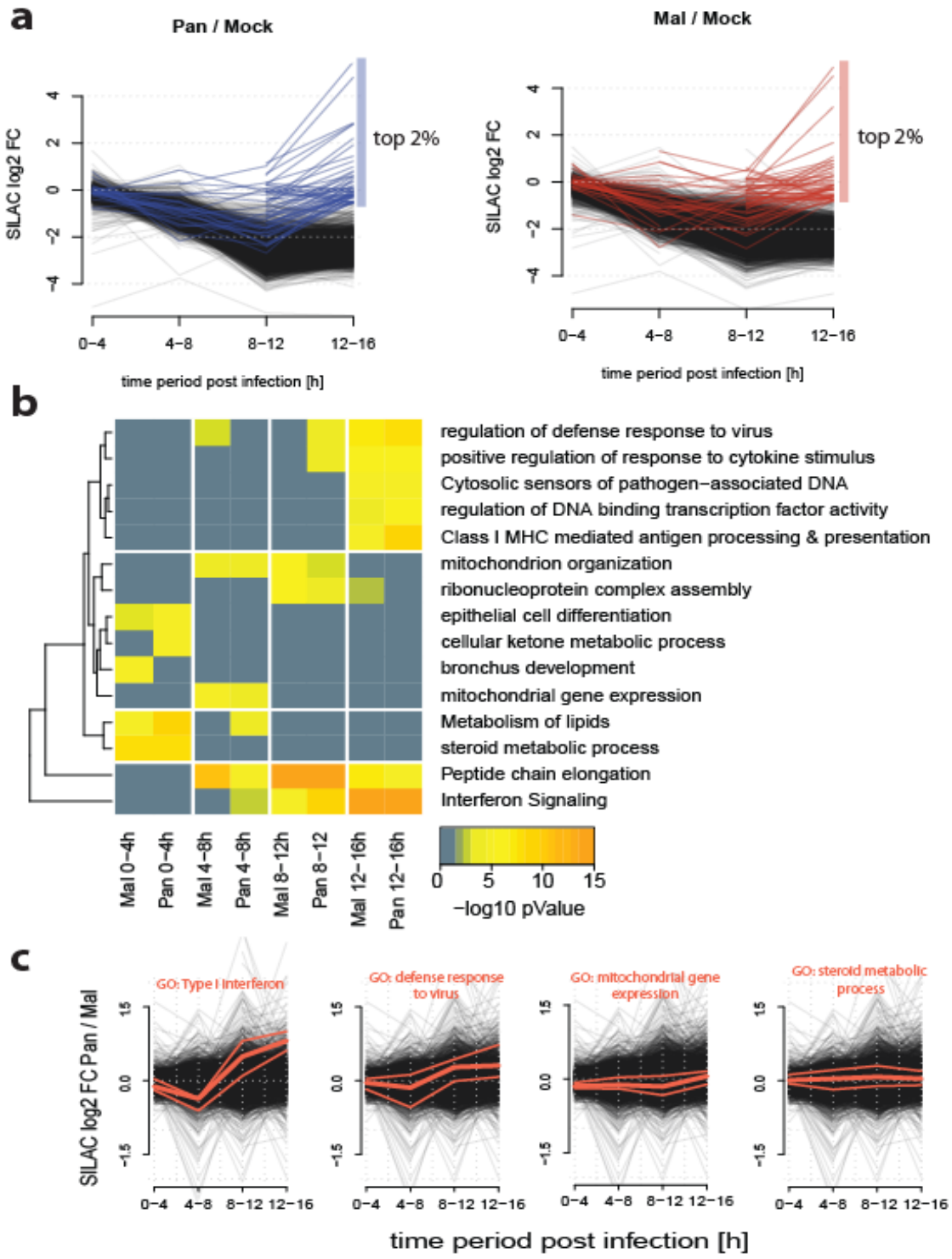


Figure 11 The dynamic host proteome of IAV infected cells.

(a) *SILAC profiles of host proteins infected with Pan (left panel) or Mal (right panel). The top 2 % of proteins that are strongest produced during the last time period are highlighted.* ***(b)*** *Functional protein clusters least affected by shutoff were identified. The two percent proteins with the highest ratios were selected for each time period and multi-list GO enrichment was performed.* ***(c)*** *Median, 25th and 75th percentile of proteins related to a selected set of GO terms are highlighted red in the direct SILAC comparison Pan / Mal. All other protein profiles in black. All data are based on the average of two replicates.*

5.2.3. mRNA levels determine IAV induced host shutoff

Several mechanistic explanations were put forward as to why IAV selectively produces viral proteins over host proteins. This includes mechanisms at the transcriptional (Rodriguez et al., 2007), post-transcriptional (Bercovich-Kinori et al., 2016; Jagger et al., 2012) and translational level (Garfinkel and Katze, 1992, 1993). To gain mechanistic understanding of the IAV induced host shutoff, we quantified mRNA levels at 8 hours post infection by RNA sequencing. We then assessed how mRNA fold-changes correlate with protein synthesis level fold-changes across all time intervals. As expected, mRNA level differences at this time point showed relatively good correlation with corresponding differences in *de novo* protein synthesis, particularly during the subsequent 8-12 h period (Figure 12a,b). Interestingly, the correlation coefficient for Pan infection was lower than for Mal infection (0.34 vs 0.59). This may indicate that additional post-transcriptional processing mechanisms are in place during Pan infection. Collectively, we observed that mRNA level changes play an important role for the shutoff of individual mRNAs. This finding corroborates the view that the host shutoff is mainly due to reduced host mRNA levels (Bercovich-Kinori et al., 2016; Inglis, 1982).

Next, we addressed whether viral mRNAs are translated more efficiently than host mRNAs. We investigated this by calculating protein synthesis efficiencies (i.e. the amount of protein made per mRNA). Therefore, we divided iBAQ values by corresponding RPKM values (Figure 12c). Upon infection with either strain host protein synthesis efficiencies were reduced compared to uninfected controls. Surprisingly, viral proteins were synthesized less efficiently than host proteins in both

strains. This suggests that mRNAs from human and avian Influenza virus strains are similarly translated. Also, this argues against the possibility that modulation of translation efficiency is related to species specificity.

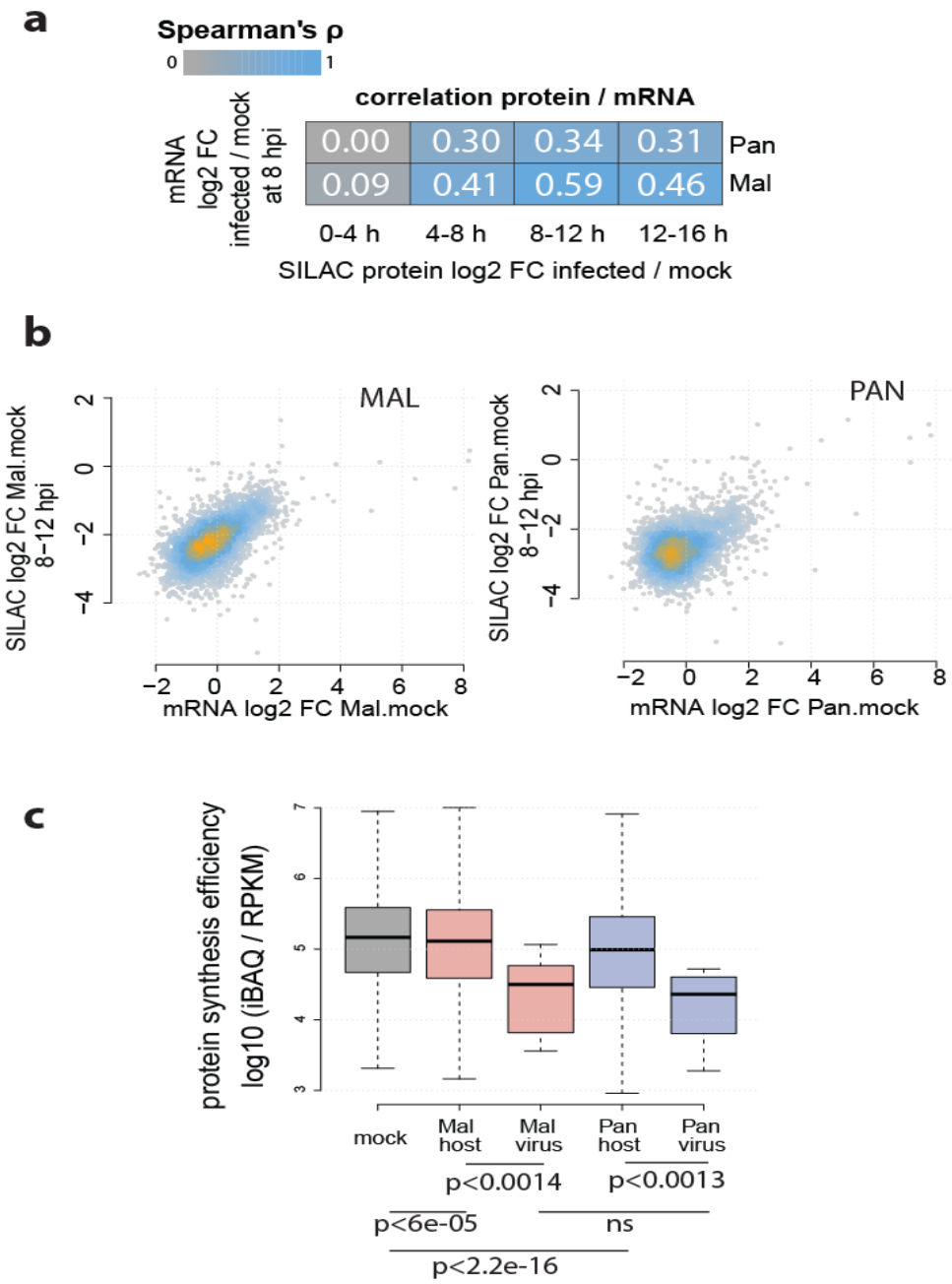


Figure 12 Cross-comparison of RNA and protein level data.

(a) *Correlation coefficients when comparing the indicated RNA and protein level data.* ***(b)*** *Scatterplot of mRNA changes versus protein synthesis level changes for the different viruses, as indicated. Blue and orange coloring of data points reflects the density of data points.* ***(c)*** *Protein synthesis efficiencies calculated from absolute mRNA (RPKM, 8 hpi) and protein data (iBAQ, 8-12 hpi). Data points outside the whiskers were removed for visibility. For the assessment of statistical significance two-sided wilcoxon rank sum tests were performed and p-values are given (ns: non-significant). All data are based on the mean of two replicates at both, the mRNA and protein level. The RNAseq data that were used for the analysis were obtained by Xi Wang, Jingyi Hou and Wei Chen.*

5.3. The dynamic viral proteome

5.3.1. Kinetics of viral protein synthesis

The differences in the synthesis of host proteins comparing both strains were relatively small. This prompted us to investigate viral protein synthesis. We first looked at iBAQ ratios to estimate the absolute levels of viral protein synthesis (Figure 13). We found that the synthesis of most proteins peaked during the 8-12 h interval. Influenza A viral proteins are categorized based on their expression kinetics as early or late. We observed the synthesis of some viral proteins, such as NP and polymerase components, already during the first pulse interval. The iBAQ values of other proteins such as HA and M1 were first recorded during the 4-8 h interval. Based on whether the iBAQ values are recorded during the first or second time interval, we categorize the proteins NP, NS1, PA, PB1, PB2 as early and M1, M2, NS2, HA and NA as late viral antigens. This recapitulates the known kinetics of early and late viral protein synthesis (Shapiro et al., 1987).

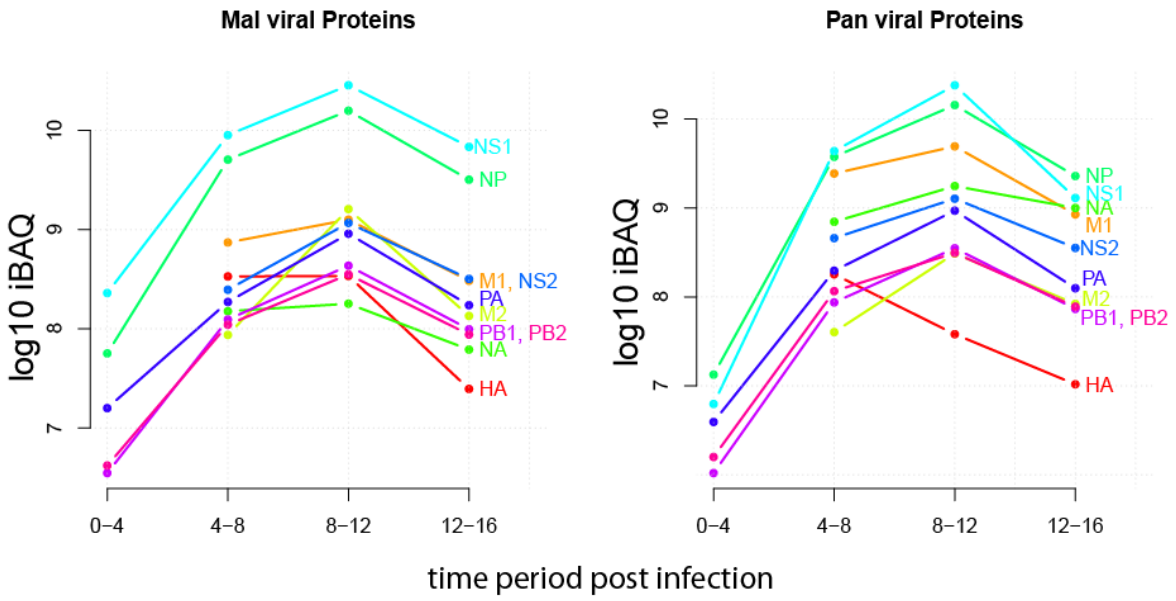


Figure 13 Viral protein synthesis kinetics.

iBAQ-based absolute quantification of protein synthesis across infection for the indicated 10 viral proteins of strain Mal (left) or strain Pan (right). The average of n=2 replicates is depicted.

5.3.2. Dysregulated synthesis of viral proteins during non-permissive infection

To accurately compare the synthesis of viral proteins between both strains, we quantified based on shared peptides (that is, peptides with sequence identity between both strains) (Figure 14a,b). We found that the avian strain produced higher amounts of all viral proteins at the beginning, indicating that the Mal virus enters cells successfully and initiates its gene expression program. However, during mid to late phases, the human-adapted Pan virus produced most proteins more abundantly than the avian strain. Kinetic profiles for NS1 and M2 proteins could not be recorded as no shared peptides were identified.

One important and well-established factor that determines host range is the RdRp. Typically, avian-adapted polymerases are less active in mammalian cells than polymerases from mammalian-adapted strains (see 1.1.4.5. Molecular determinants of species specificity and (Gabriel et al., 2005; Long et al., 2016; Long et al., 2019)). We thus would have expected that synthesis of all viral proteins is similarly attenuated with the avian strain. However, we observed striking differences in the synthesis of individual proteins. The HA protein was produced stronger by the avian strain during the entire course of infection. In contrast, the Matrix protein M1 and NA were more abundantly produced by the human strain, especially at later stages. It is unlikely that these differences arise due to global differences in the polymerase activity between both strains. Thus, viral protein production during a non-permissive infection is dysregulated.

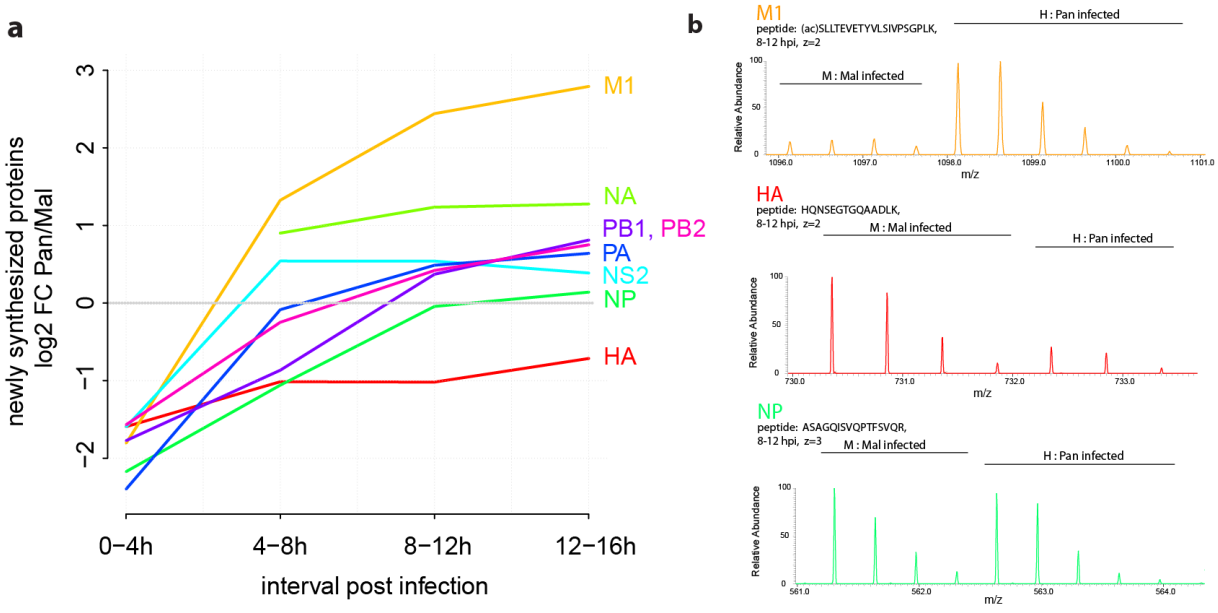


Figure 14 Viral protein synthesis is dysregulated during non-permissive infection.

(a) *SILAC protein synthesis profiles of human vs avian IAV proteins based on shared peptides. No shared peptides were identified for NS1 and M2. The average of n=2 replicates is depicted.* ***(b)*** *MS1 spectra of individual precursor (M1 - top, HA - middle, NP - lower) peptides in M and H SILAC channels.*

5.3.3. NP is inefficiently exported during non-permissive infection

We then planned to assess the regulation of one or more proteins in more detail. First, we tried to functionally correlate the differences in synthesis levels of individual proteins with functional infection parameters. First, we focused on the HA protein, the only protein which was stronger produced by the avian-adapted virus throughout the infection. It was previously shown that membrane accumulation of HA in infected cells can promote MAPK (mitogen activated protein kinase) signaling (Marjuki et al., 2006). We therefore assessed whether MAPK targets are differentially expressed in A549 cells infected with Pan or Mal strains in our data using recently published data on MAPK-responsive genes in A549 cells (Yue et al., 2017). Specifically, we assessed how genes that are MAPK transcriptionally responsive and genes that are transcriptionally inert to MAPK ("control genes", (Yue et al., 2017)) are expressed during Pan or Mal infection (Figure 15a). We observed that MAPK responsive genes were stronger expressed under Pan than Mal infection. Hence, this analysis does not support increased HA-induced MAPK activation by the Mal strain relative to Pan.

Then we concentrated on a different protein, M1. This has the following reasons: i) It is well conserved between both strains at the amino acid level (~96 % amino acid identity). ii) It showed the highest fold-change in infected A549 comparing both strains. iii) It is the most abundant protein in virions across many IAV isolates (Hutchinson et al., 2014) and iv) It is an essential viral protein that has many functions during infection, including the export of vRNPs from the nucleus to the cytosol (Bui et al., 2000; Martin and Helenius, 1991a). In fact, M1 expression at late stages of infection is essential for the accumulation of vRNPs in the cytosol of infected cells.

To assess whether the export of vRNPs is disturbed during a non-permissive infection, we investigated the subcellular distribution of the viral nucleoprotein by immunofluorescence microscopy. While the Pan strain (Figure 15b,c) exported NP efficiently to the cytoplasm, NP accumulated in the nucleus upon Mal infection. This observation is independently corroborated by the fact that Mal infection elicits a weaker interferon response at late stages than the human strain (Figure 11c). The stimulation of the interferon response by viral RNAs in the cytosol (Killip et al., 2015) may thus be impaired with the avian strain. In conclusion, the infection with the avian-adapted virus correlates with reduced M1 synthesis and nuclear retention of NP in human cells.

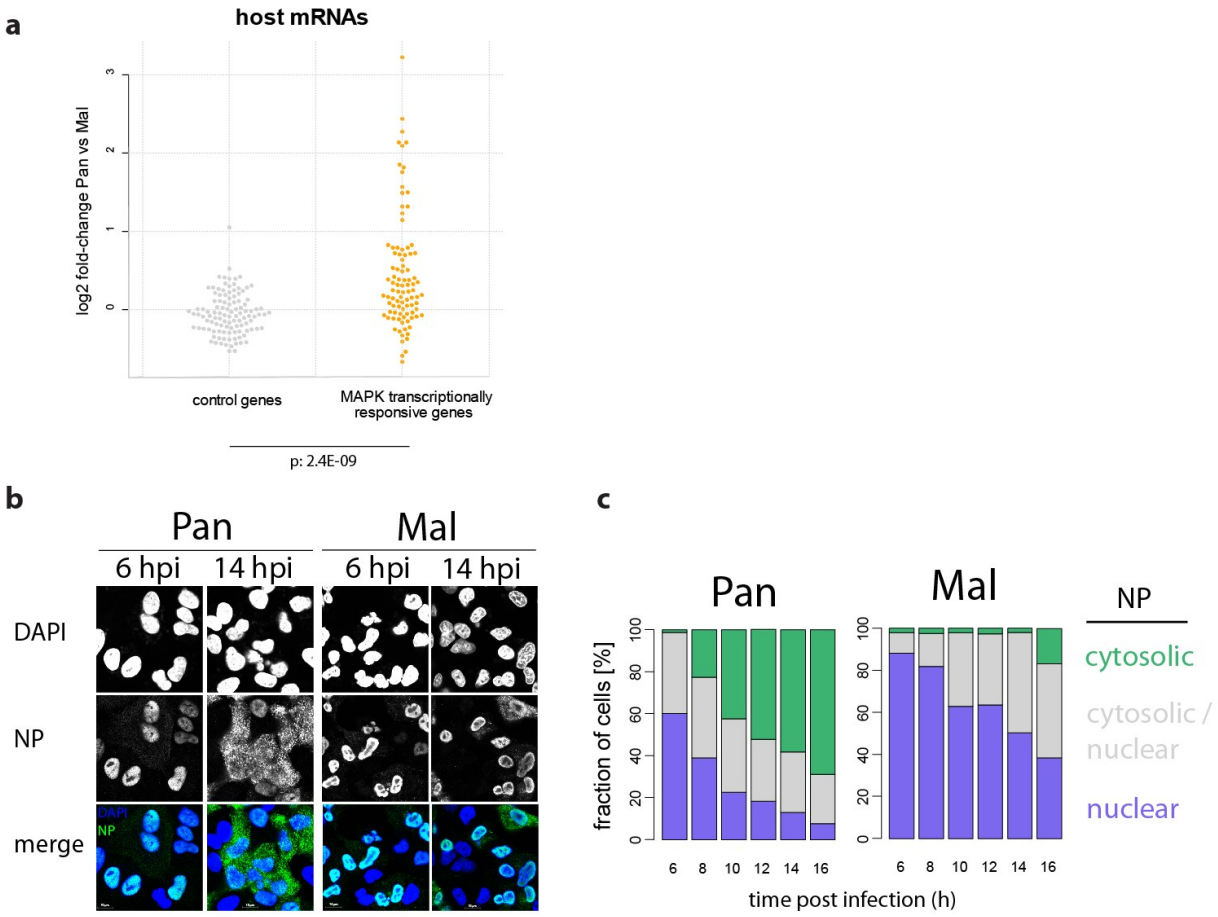


Figure 15 Functional correlation of protein synthesis data with infection parameters.

(a) *Relative mRNA levels of MAPK responsive genes (Yue et al., Genes Dev, 2017) and control genes comparing Pan and Mal infection (p-value based on two-sided Mann-Whitney test). The average of n=2 replicates is depicted.* ***(b)*** *Cells were infected with the indicated viruses at MOI=1 (FFU/cell) and analyzed by immunofluorescence microscopy for NP trafficking. Nuclei were counterstained with DAPI.* ***(c)*** *For quantification, NP staining pattern was categorized as predominantly cytosolic, nuclear or both. At least 150 cells were counted per condition. The RNAseq data depicted in panel a is based on RNA sequencing performed by Xi Wang, Jingyi Hou and Wei Chen.*

5.4. Differences in the control of IAV RNA splicing

5.4.1. Dysregulated splicing of M and NS segment RNA

We next asked about the reasons for the low level synthesis of M1. We hypothesized that mRNA levels are determining the levels in protein synthesis. To address this hypothesis, we quantified the levels of viral mRNAs from RNA-sequencing data (Figure 16a). We observed that the avian virus produced ~2/3 of the mRNA of the human strain with the single largest difference for M1. Also, the strain-specific differences in the mRNA levels for M1 reflect the strain-specific differences at the protein level. This indicates that the reduced levels for M1 protein are caused by low M1 mRNA levels (Figure 16b). Differences in the strain-specific production of other proteins such as HA and M1 can also be explained by differences in mRNA abundances.

The M1 protein is produced from a collinear M1 transcript that can be further spliced to several isoforms (see 1.1.4.2. IAV RNA splicing). We quantified the proportion of these transcripts in the RNA-sequencing experiment based on splice junction reads (Figure 16c,d). We observed that the avian strain produces all known transcripts, plus a novel isoform that we call RNA 5. This transcript is a spliced isoforms with the 5' GG site (pos 520/521) as the donor and the common 3' acceptor site. It contains an ORF in-frame with M1 with a missing internal region. While, we observed that the avian strain splices about 27 % of its RNAs to alternative isoforms, the human strain splices only a few percent (2-3 %). This indicates that the relatively low levels of the M1 mRNA are at least partially due to increased M segment RNA splicing with the avian strain.

To validate this finding we measured the kinetics of M1 and M2 mRNA abundances by qRT-PCR (Figure 16e). This confirmed that M1 was weakly expressed with the avian strain. In contrast, we observed that the absolute M2 levels were more similar comparing both strains. There are two opposing effects that result in comparable M2 levels: On the one hand, the increased splicing of the collinear transcript. On the other hand, the global reduction in overall transcript levels, which results probably from the low activity of avian-adapted polymerase in human cells (Gabriel et al., 2005). In conclusion, M segment RNA splicing is dysregulated for the avian virus in human cells.

IAV contains another segment that can be alternatively spliced (that is the NS segment or segment 8). To assess whether NS segment RNA splicing is also dysregulated during infection with the Mal strain, we quantified RNA isoforms originating from the NS segment based on splice junction reads (Supplementary Figure 3a,b).

Similar to the M segment, the NS segment encodes for a collinear transcript that can be spliced to alternative isoforms (NS3 and NS2 mRNAs). Splicing to the NS3 isoform was previously linked to human adaptation of IAV (Selman et al., 2012). Consistently, we observed splice junction reads for the NS3 isoform exclusively for the human-adapted IAV strain. In addition, about 13 % of the primary NS transcript is spliced to NS2 for the human-adapted strain. In contrast, the avian-adapted strain splices only 5 % to NS2. Thus, splicing of the NS segment is dysregulated with the avian strain in mammalian cells. Importantly, the extent to which splicing is dysregulated is more pronounced for the M segment.

Taken together, non-permissive infection with the avian IAV isolates correlates with dysregulated splicing of NS and M segment RNAs. Also, the low levels of M1 mRNA

during non-permissive infection can at least partially be explained by high levels of M segment RNA splicing.

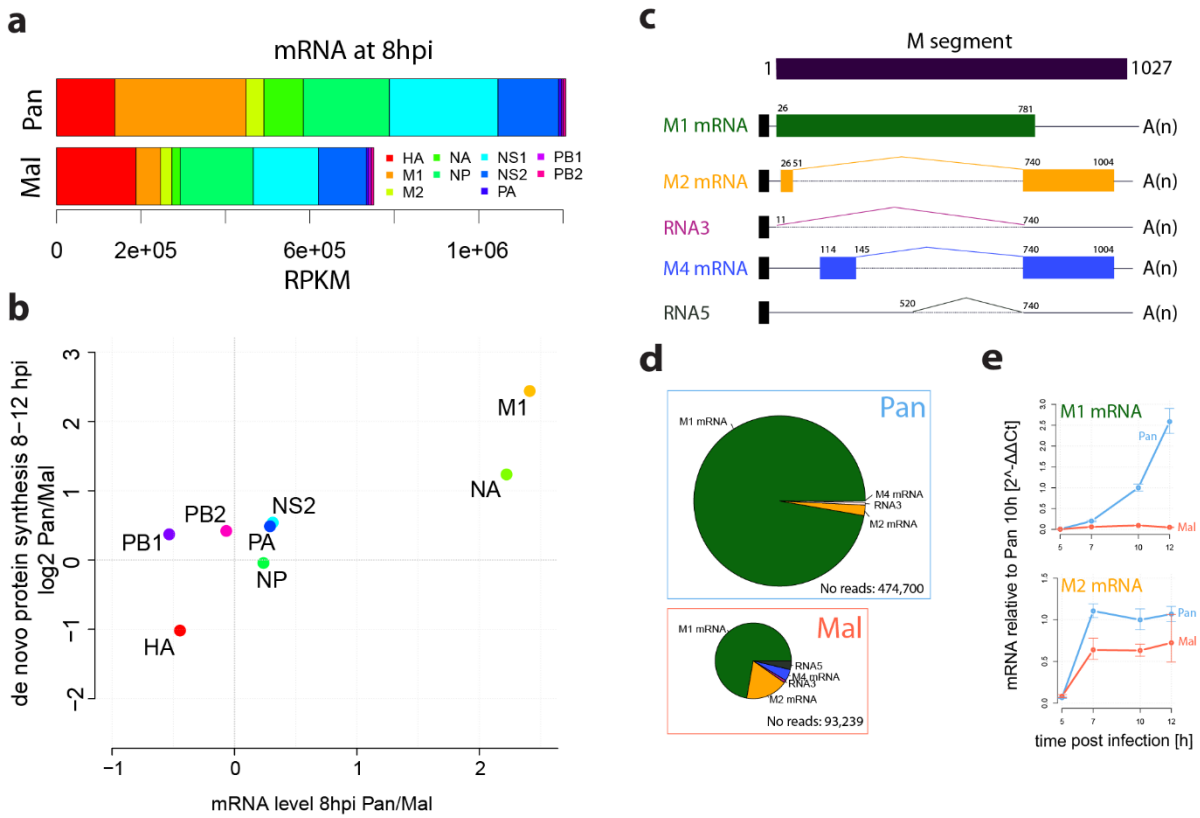


Figure 16 M segment RNA splicing is markedly different in permissive versus non-permissive infection.

(a) *Normalized read counts (RPKM) for 10 viral mRNA from poly-A enriched samples from either Pan or Mal infected samples at 8 hpi. Data represent means of two independent experiments.* ***(b)*** *Direct comparison of fold-changes for viral mRNAs/ proteins from RNA level and protein synthesis data. Data represent means of two independent experiments.* ***(c)*** *Schematic depiction of M gene architecture. The M pre-mRNA can be spliced into the indicated isoforms.* ***(d)*** *Relative quantification of the different isoforms based on splice junction reads from RNAseq data for both strains at 8 hpi. The area of the pies reflects the absolute number of splice junction reads Data represent means of two independent experiments.* ***(e)*** *M1/M2 mRNA abundance kinetics based on qRT-PCR of cells infected with the indicated viruses at MOI=4 (FFU/cell). One experiment is shown with triplicate measurements. The RNAseq data that were used for the analysis were obtained by Xi Wang, Jingyi Hou and Wei Chen. Further, Xi Wang performed the analysis on splice junction reads. Immanuel Husic performed the qRT-PCR measurements.*

5.4.2. The 3’ splice site region controls M segment RNA splicing

In principle, the observed differences in M segment RNA splicing can be caused by numerous factors. This includes i) *trans*-acting factors (that is, viral or host proteins that are present during infection), ii) *cis*-acting factors (that is, sequences in the M segment such as splicing enhancer or silencer elements as well as RNA secondary structures), or iii) a combination of both.

First, we aimed to understand whether there is *cis*-acting control over M segment RNA splicing. To address this possibility we designed the following reporter system: We cloned the coding region of the M gene into a eukaryotic expression vector under the control of a CMV-promoter and fused it to an N-terminal Flag/HA-tag (Figure 17a). Cells transfected with these constructs produce two Flag/HA-tagged polypeptides. One unspliced Flag/HA-M1 with relatively high molecular weight and one spliced Flag/HA-M2 with lower molecular weight. When we transfected wild type constructs into human A549 cells, we detected high levels of the spliced product M2 with the avian construct. In contrast, cells transfected with the human construct produced only very little M2 (Figure 17b,c). Thus, the reporter system reflects the splicing differences during infection and indicates that *cis*-acting signals are involved in the strain-specific control of splicing.

To determine the sequence responsible for the strain-specific splicing we created chimeric human/ avian reporter constructs. When we swapped the entire intron sequence of the M2 isoform (nucleotides 52-739, which corresponds to ~70 % of the coding sequence), we did not observe major changes in the relative levels of M1 to M2. In contrast, when we inserted the human 3’ splice site region (nucleotides 707-

779, 8 nucleotide differences between Pan and Mal) into the avian backbone, we observed that splicing was impaired down to the levels of the human wild type construct. Likewise, integration of the avian 3' splice site resulted in a strong increase in splicing, mimicking the avian wild type construct. We validated these results at the mRNA level employing qRT-PCR and found that the 3' splice site region alone is sufficient to switch the species specific splicing phenotype. Thus, the 3' splice site region encompasses a *cis*-acting regulatory element that controls the species-specific M segment splicing pattern. An alignment of Pan and Mal M segment nucleotide sequences containing the coordinates of splice donor and acceptor sites as well as the 3' splice site region is given in Supplementary Figure 4. The RNA in the 3' splice site region was previously shown to contain RNA secondary structure (Moss et al., 2012) and to bind to the ASF/SF2 (SRSF1) splicing factor (Shih and Krug, 1996).

The RNA polymerase II (pol II) is known to influence splicing patterns due to the association of transcription with pre-mRNA processing (Hicks et al., 2006). The reporter system employed here relies on transcription of the cellular pol II, which may influence splicing in other ways than an IAV RdRp. Thus, we aimed to validate our findings in a reporter system employing the IAV RdRp. Therefore, we used pHW2000 plasmids encoding the NP, PA, PB1 and PB2 segments that produce when transfected into HEK293T cells a replication and transcription active IAV RdRp (Supplementary Figure 5a). When a pHW2000 plasmid encoding the M segment was added in addition, M1 protein accumulated (Supplementary Figure 5b). Upon, omission of the pHW2000 plasmid encoding NP from the transfection setup M1, was not detected, indicating that accumulation of M1 in this system predominantly depends

on the activity of the viral polymerase. Then we assessed the levels of M1 and M2 proteins with this system using wild type and chimeric Mal/ Pan M segments (Supplementary Figure 5c). Upon transfection of the chimeric human M segment containing the 3' splice site region (707-779) of the Mal strain (termed Pan-Av for Panama with avian 3' splice site region) we detected lower M1 but higher M2 levels relative to the transfection with the Panama wild type construct. Likewise, upon transfection of the avian chimeric M segment containing the Pan 3' splice site region, (Mal-Hu, for "Mal" with a human 3' splice site) M1 levels were higher and M2 levels were lower relative to the wild type avian construct. This result confirms that the 3' splice site region is a *cis*-regulatory element controlling M segment RNA splicing.

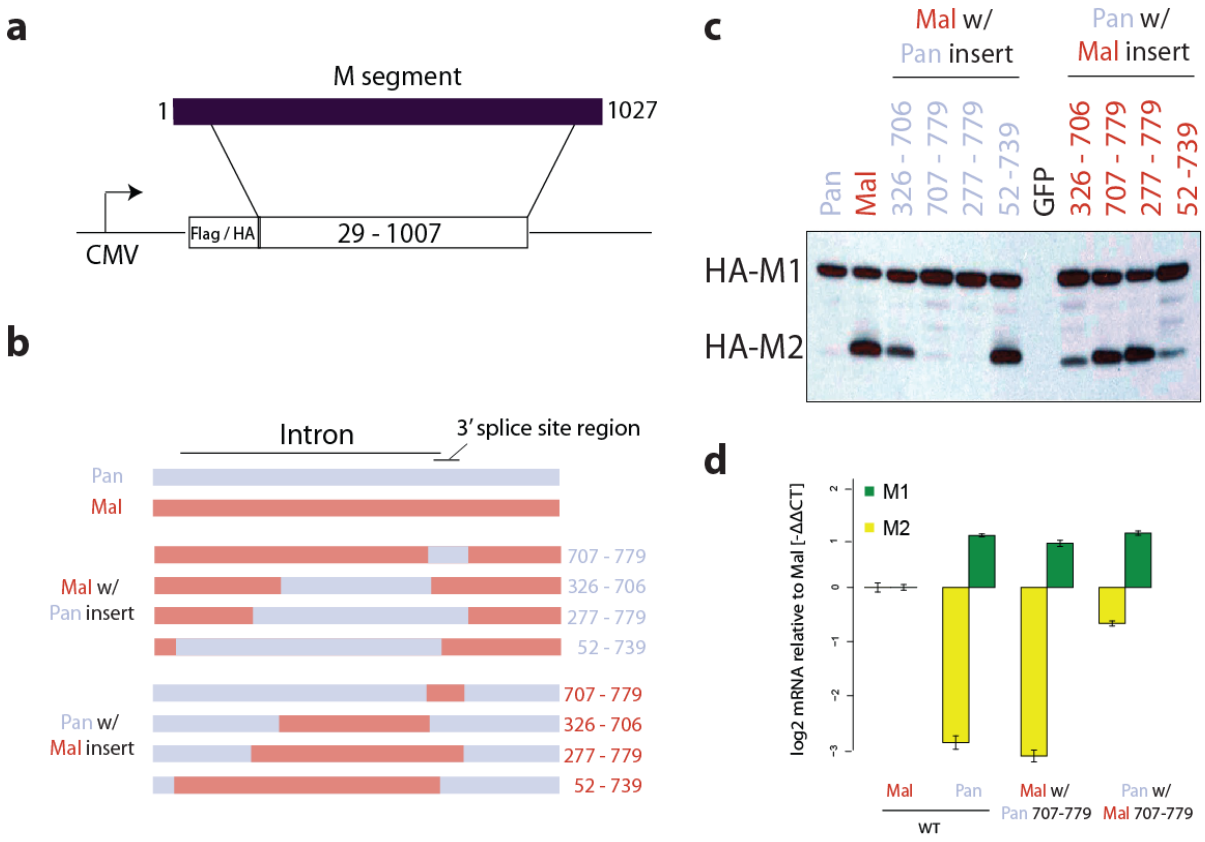


Figure 17 A cis-regulatory sequence element in the 3' splice site region controls M segment RNA splicing efficiency.

(a) *Reporter system. The coding sequence of segment 7 was cloned into a eukaryotic expression vector with N-terminal Flag/HA tag.* ***(b)*** *Wild type Pan and Mal sequences as well as several chimeric Pan/Mal constructs were cloned. A549 cells were transfected with the indicated reporter constructs, harvested and subjected to anti-HA immunoblotting* ***(c)*** *or qRT-PCR* ***(d)****. The qRT-PCR data is based on single experiment with triplicate measurements. The qRT-PCR measurements were performed by Immanuel Husic. The plasmids for the chimeric constructs depicted in panel b were created in collaboration with Barbara Vetter and Lüder Wiebusch.*

3.5.4. Differences in RNA secondary structures at the 3'splice site

We next wanted to address how relevant the control via the 3'splice site is for other IAV strains. In particular, we were interested in identifying differences in RNA secondary structures that were conserved for avian and human-adapted IAVs during their evolution. For this reason we teamed up with Irmtraud Meyer (MDC Berlin) who predicts RNA secondary structures bioinformatically. She performed multiple sequence alignments of hundreds of avian and human H3N2 M segment RNAs and then performed RNA structure analyses with the program RNA decoder (Pedersen et al., 2004). This program was previously successfully applied to identify the evolutionarily conserved structures in viral genomes, such as HCV and HIV (Pedersen et al., 2004; Watts et al., 2009). Importantly, it is capable of dis-entangling the overlapping evolutionary constraints at the amino acid and RNA level. In contrast to other algorithms for RNA structure prediction, the algorithm is not based on calculations of thermodynamic stability *in vitro* but instead only uses the aligned sequences provided as input. Further, it can estimate the reliability of its own predictions by probability scoring.

When using this algorithm, we found that the M segment RNA structure that is best supported is markedly different between human and avian IAVs (Figure 18a). In particular, the region at the 3' splice site contains two species-specific RNA secondary structures. The avian region contains a hairpin that overlaps the 3'splice site (Figure 18b). This structure is similar to a hairpin reported by Moss et al., (Moss et al., 2012) for four avian strains. A side-by-side comparison of the avian-predicted structure in this thesis with the two structures proposed by Moss et al. is given in Supplementary

Figure 6. In contrast, the human structure involves a stem, bulge and a terminal loop and is located 3' to the splice site involving positions in the second exon of the M2 splice variant only (Supplementary Figure 7). The predicted structures predominating in avian or human-adapted viruses are mutually exclusive.

To validate the existence of different RNA secondary structures, we first *in vitro* transcribed RNAs in the region that is proposed to be structured for both strains (707-825, 14 nucleotides different between Pan and Mal). Then we purified the RNAs, refolded them and ran them on native and denaturing agarose gels. We observed that Pan and Mal RNAs had different migration patterns in native gels, indicative of differences in RNA secondary structure (Figure 18c). When we mutated 8 nucleotides at the 3' splice site region in the Mal RNA to the corresponding Pan sequence, this resulted in a migration pattern similar to the wild type Pan RNA. Conversely, mutating 8 nucleotides in the Pan RNA to the sequence of the Mal strain resulted in a similar RNA structure as with the wild type Mal RNA. All RNAs migrated in single bands at the same molecular weight in denaturing gels. Thus, the M segments of avian and human IAV strains display different conserved RNA secondary structures exactly in the region relevant for strain-specific control of M segment RNA splicing.

5.4.3. Analysis of other IAV isolates

In addition to the experiments about RNA secondary structures, we also wanted to test the functional relevance of our findings for other IAV isolates experimentally. Therefore, we cloned several M segments of human, avian and swine IAV origin into our reporter system (Figure 18d,e). The M segment of the seasonal Pan strain stems

from the M segment of the A/Brevig Mission/1/1918 (p1918) virus. This virus is the founding virus of currently circulating human strains and caused the 1918 “Spanish flu” pandemic (Worobey et al., 2014a) (see 1.1.4.3. IAV epidemics and pandemics). When A549 cells were transfected with the p1918 M segment reporter vector (Figure 18e), we observed inefficient splicing of the p1918 M gene, consistent with our data for the Pan strain and previous reports (Backstrom Winquist et al., 2012). Mutating the 3’ splice site region to the sequence of the Mal strain increased splicing. In addition, we cloned the M segment of A/Rostock/Germany/45/1934 (fowl plague virus – FPV), a virus that is at the evolutionary root of eastern hemisphere AIVs (Worobey et al., 2014b). We observed that this M segment is spliced very efficiently (Figure 18d, lane 3). Strikingly, integrating the 3’ splice site of the Pan strain led to very inefficient splicing (6 nucleotides mutated, Figure 18d, lane 4). Thus, while inefficient splicing of the M gene occurs in human-adapted IAVs (seasonal: Pan and pandemic: p1918), efficient splicing occurs in avian-adapted IAVs (Mal and FPV).

Occasionally, IAV may cross the species barrier via intermediate hosts such as swine (Neumann et al., 2009). To test how M segments of swine-origin IAV splice the M segment, we cloned the M segment of A/swine/Netherlands/25/1980, which is considered one of the earliest “avian-like” swine-origin IAVs (Smith et al., 2009). We observed that this segment is spliced similarly as that of the Mal strain, suggesting that it is insufficiently human-adapted (Figure 18d, lane 2). In contrast, splicing of the M segment of a highly pathogenic IAV of H5N1 subtype (A/Vietnam/1203/2004) that originated from a fatal case of IAV in humans (Tran et al., 2004), was relatively inefficient (Figure 18d, lane 5). This suggests that this M segment shows signatures

of adaptation to the human host, which may contribute to the pathogenicity of A/Vietnam/1203/2004 in humans.

Collectively, our results indicate that avian and human viruses splice the M segment with different efficiencies, which can be modulated by the 3' splice site region. Further, these result support the global importance of M segment splicing with respect to host range.

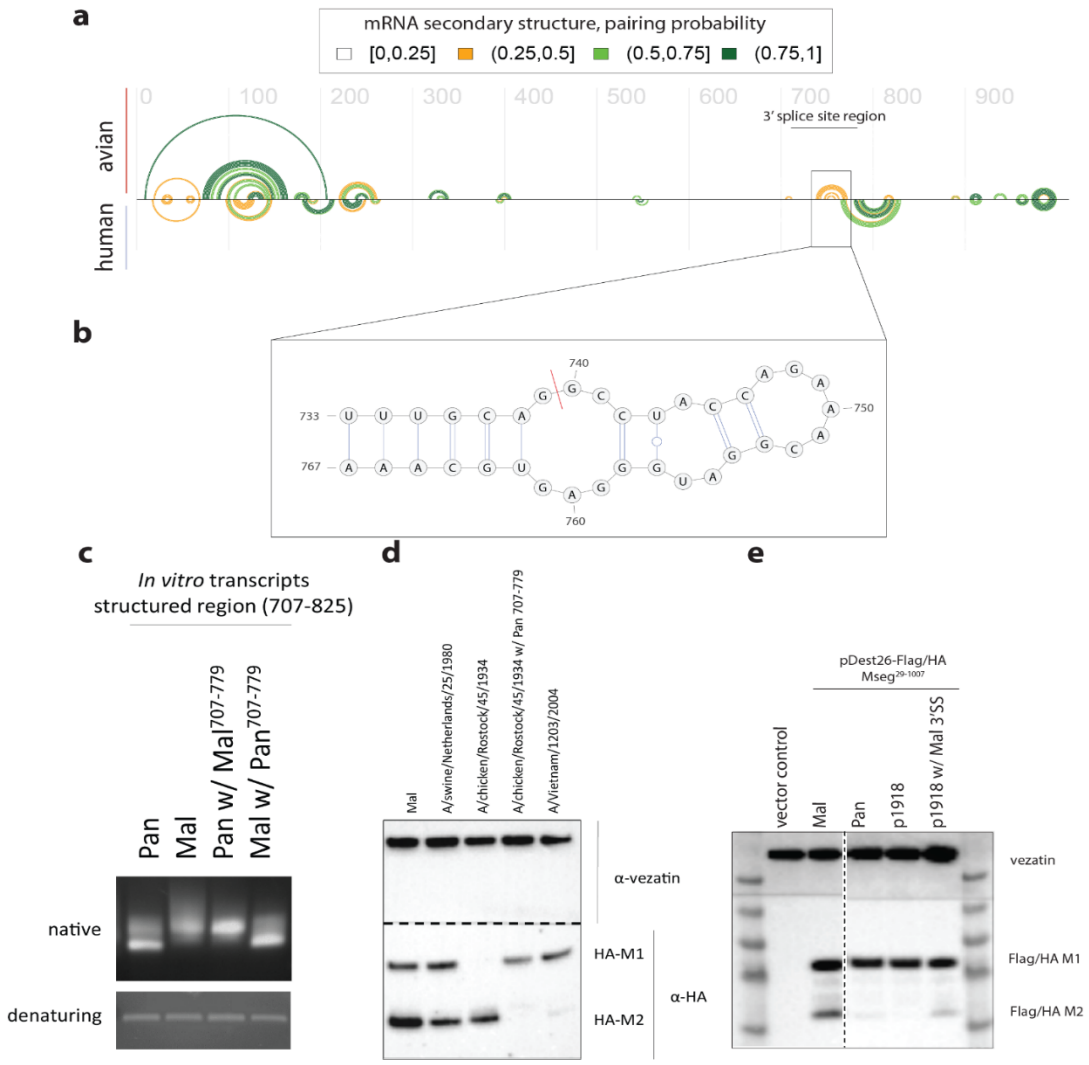


Figure 18 Conserved differences in RNA secondary structure at the 3' splice site.

(a) *Evolutionary conserved RNA secondary structure along the M segment primary transcript predicted for avian adapted (top) and human adapted (bottom) IAV strains. Each semi-circle corresponds to one base-pair involving the corresponding two alignment positions. Color-coding of base-pairs according to their corresponding, estimated base-pairing probability. The 3' splice site region is annotated with a bar.* ***(b)*** *The insert shows a detail of the predicted RNA secondary structure at the 3' splice site, predicted for avian but not human sequences in this study. The splice position is indicated by a red bar.* ***(c)*** *The region that is predicted to be structured with two species-specific mutually exclusive secondary structures (707-825) was in vitro transcribed. The RNA was purified and subjected to native and denaturing agarose gel electrophoresis (agarose gels stained with ethidiumbromide).* ***(d,e)*** *M segments of the indicated avian, swine, human or mutant strains were cloned into the M segment splice reporter construct (pDEST 26 – Flag/HA). A549 cells were then transfected with the indicated constructs and subjected to lysis and immunoblotting. The analysis depicted in panel a is courtesy of Irmtraud Meyer. Further, Martha Hergeselle contributed to cloning the constructs depicted in panels d and e.*

5.4.4. The 3'splice site region is important for permissive infection

The experiments using reporter constructs described above are useful as they allow us to study the impact of M segment sequence features independent of a viral infection. However, it is critical to validate and test the importance of our findings also during infection. Therefore we mutated the Pan M segment 3'splice site towards the sequence of the corresponding Mal strain (8 nucleotides mutated) using reverse genetics ("Pan-Av" for Pan virus with an avian splice site) (Figure 19a).

First, we compared the kinetics of viral protein production between both mutant and wild type strains using pulsed SILAC (pSILAC) in A549 cells (Schwanhausser et al., 2009). During both pulse intervals (6-12 hpi, 12-18 hpi), M1 synthesis was selectively impaired. At later stages, the Pan-Av strain also presented with a decrease in the production of other essential viral proteins (Figure 19b), probably reflecting impaired viral replication. Next, we quantified M1 and M2 protein (Figure 19c) and mRNA levels (Figure 19d). M1 protein /mRNA levels were decreased with the Pan-AV mutant relative to the wild type Pan virus. This behavior is similar to that of the Mal strain (compare also Figure 14). Then we planned to assess the importance of the avian 3' splice site region in the human-adapted IAV background for multiplication of the virus. Therefore we assessed the multiplication characteristics of the different viruses (Figure 19e). We observed that, as expected, the avian-adapted virus produces much fewer novel progeny than the human-adapted virus (~1,000 fold fewer). The exchange of the entire M segment of the Pan with that of the Mal strain caused about ~10 fold attenuation in viral titers (Pan + Mal M). A similar ~10 fold attenuation was seen with

the Pan-Av mutant virus. Thus, the human-adapted 3’ splice site of the IAV M segment is indeed important for permissive infection.

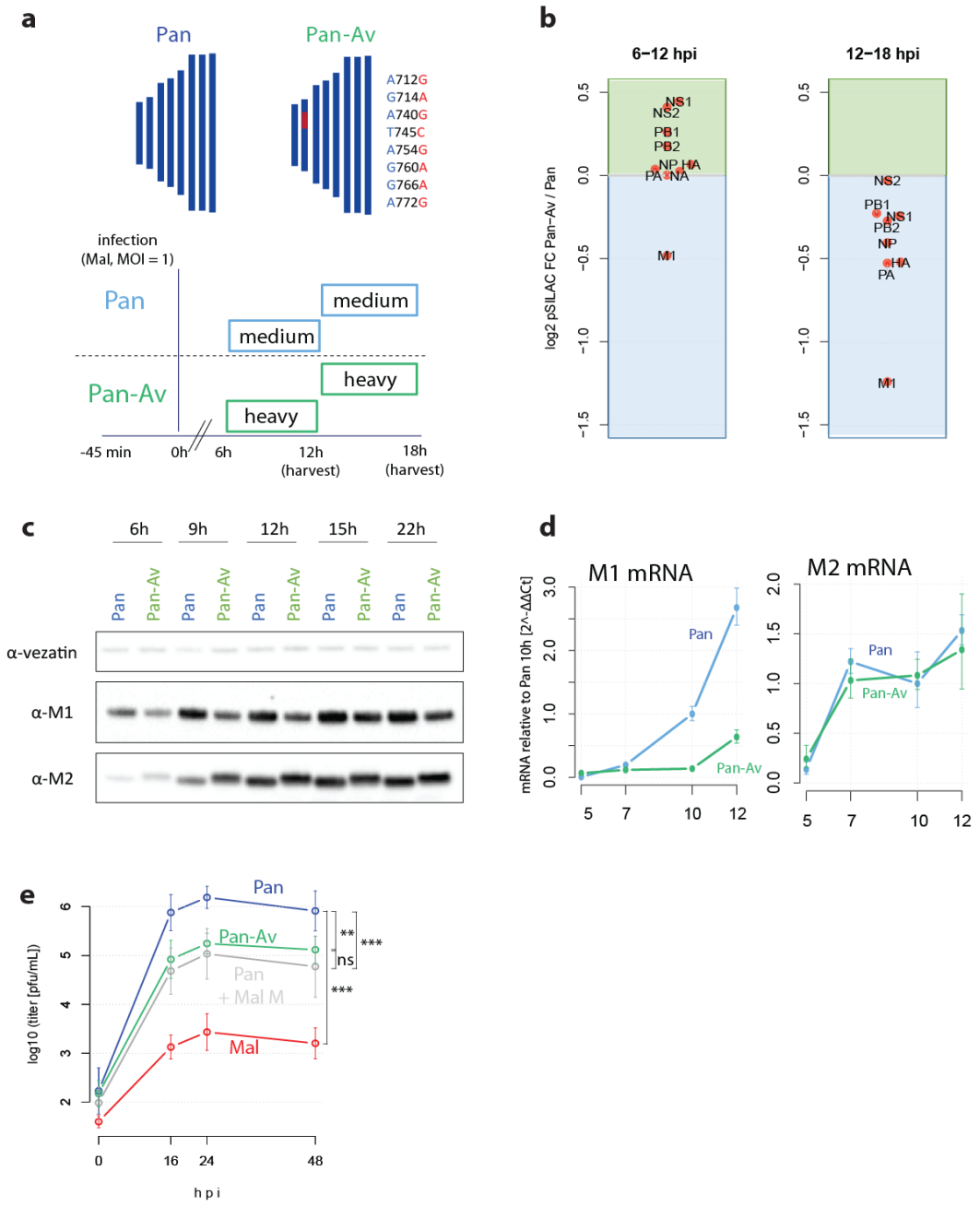


Figure 19 The 3' splice site is important for permissive IAV infection.

(a, top) *8 nucleotide polymorphisms of Mal virus at the segment 7 3' splice site were integrated into wild type Pan virus generating "Pan-Av" mutant.* ***(a, bottom)*** *Design of pulsed SILAC experiment. A549 cells were infected with wild type or mutant virus at an MOI of 4 (FFU/cell) and viral gene expression was assessed by pulse labeling during two time intervals.* ***(b)*** *SILAC-based quantification of viral protein expression comparatively for Pan and Pan-Av virus. Kinetics of M1/M2 expression was detected by immunoblotting* ***(c)*** *or detected by qRT-PCR* ***(d)****.* ***(e)*** *Multicycle replication curve of the indicated viruses at an MOI of 0.05 (FFU/cell) on A549 cells. Means and standard deviations of biological triplicates are shown along with significance estimates based on two-sided paired t-tests for the 16,24 and 48 h time points (ns: non-significant, ** : p<0.01, *** : p<0.001). Figure panel e was contributed by Gudrun Heins, Immanuel Husic performed qRT-PCR depicted in panel d.*

5.4.5. Host-specific control over M segment RNA splicing

We then asked how M1 expression is regulated in avian cells, to which the Mal virus is evolutionary adapted. Specifically, it may be that avian-adapted viruses produce more M1 in avian than human cells. This would indicate that an optimal balance of viral proteins is needed irrespective of the host species. In contrast, low levels of M1 in both avian and human cells would reflect an adaptation towards a decreased need of M1 protein.

To address both possibilities, we aimed to compare viral protein production of the Mal virus in human (A549) versus avian (DF-1) cells. To do this, we pulse labeled cells during two time intervals with heavy and medium SILAC amino acids and then analyzed the SILAC ratios of viral proteins (Figure 20a,b). Again, during non-permissive infection, protein production was dysregulated with the avian strain: This effected the NA, M1 and M2 proteins most severely. Importantly, M1 is underproduced in A549 compared to DF-1 cells. This indicates that the reduced level of M1 in human cells reflects poor adaptation to the human splicing environment. In addition, this experiment identified host-specific control of M1 expression. Thus, M segment RNA splicing is not only controlled by *cis*-acting, strain-specific elements (Figure 17, Figure 18, Figure 19) but also by *trans*-acting host-specific factors.

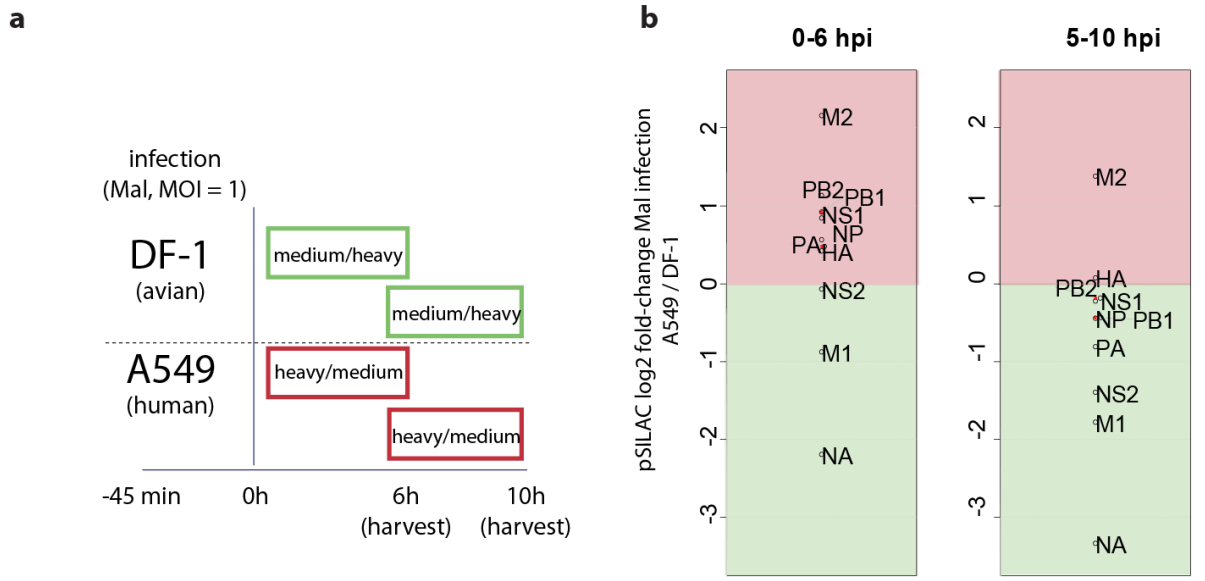


Figure 20 Host-specific control over M segment RNA splicing during non-permissive and permissive infection.

(a) *SILAC-light labeled DF-1 and A549 cells were pulse labeled with heavy or medium heavy SILAC amino acids (in label-swap duplicates) after infection with the avian (Mal) isolate.* ***(b)*** *pSILAC fold-changes are depicted for both pulse intervals comparing the production of viral proteins with the Mal strain in human versus avian cells. The average of n=2 replicates is depicted.*

5.4.6. *Trans*-acting control over M segment RNA splicing

Importantly, the viral protein NS1 was also shown to regulate M segment RNA splicing (Robb and Fodor, 2012). We thus aimed to understand how avian-adapted and human-adapted NS1 variants can impact strain-specific M segment RNA splicing.

To this end, we cloned the NS1 of Mal and Pan strain into eukaryotic expression vectors and fused them to an N-terminal myc-tag. Then, we co-transfected A549 cells with NS1 expressing plasmids and M segment reporter wild type constructs. We observed that NS1 of the Pan strain was weaker expressed than the NS1 of the Mal strain. Such strain-dependent differences in NS1 expression levels arise from NS1 binding to the CPSF complex (Kochs et al., 2007). NS1-CPSF interaction interferes

with polyadenylation of pol II transcripts, resulting in low transgene expression levels. Also, we observed that NS1 of the Mal strain was able to enhance splicing of the M segment RNA of both the avian and the human strain (Figure 21a). The NS1 of the Pan strain did not appear to regulate splicing of the M segment RNA using the reporter system.

We then asked if there are differences in NS1 protein-protein interactions that could mediate differential control of M segment RNA splicing. Therefore, we transfected NS1 expression plasmids of the avian and human strain as well as empty vectors into triple-SILAC labeled HEK293T cells (Figure 21b). We then performed affinity-purification mass-spectrometry experiments and assessed the interaction partners of NS1 compared to the empty vector control (Figure 21c). We found that NS1 precipitated with several previously described interaction partners, such as members of the CPSF complex (CPSF1-4, FIP1L1)(Nemeroff et al., 1998), CRKL (Crk-like proto-oncogene, adaptor protein)(Heikkinen et al., 2008), Staufen Double-Stranded RNA Binding Protein 2 (STAU2)(Falcon et al., 1999) , Protein Kinase R (PKR, EIF2AK2)(Tan and Katze, 1998), and PKR associated proteins (PRKRA). We identified RuvB Like AAA ATPase 1 and 2 (RUVBL1/2, also known as Pontin and Reptin), and PNPT1 (Polyribonucleotide Nucleotidyltransferase 1) as novel interactors with NS1 of either or both strains. Interestingly, while the interaction of most proteins was not different between both NS1 variants, the CPSF complex precipitated strongly with the human-adapted NS1 and RUVBL1/2 strongly precipitated with the avian-adapted NS1 (Figure 21d, Supplementary Table 3). Next, we attempted to validate the strain-specific interaction of RUVBL1 with NS1 and therefore performed anti-HA IPs with cell lysates

transfected with Pan and Mal NS1 (Figure 21e). As, expected, RUVBL1 precipitated exclusively with the avian-adapted NS1. Interestingly, RUVBL1 is a broadly acting host-factor that inhibits many viral infections (Yasunaga et al., 2014). In conclusion, the NS1 of the avian strain is a positive regulator of M segment RNA splicing. Further, it interacts with RUVBL1, which isknown viral restriction factor. It will be interesting to further investigate the role of NS1, and potentially RUVBL1/2, in the regulation of M segment RNA splicing.

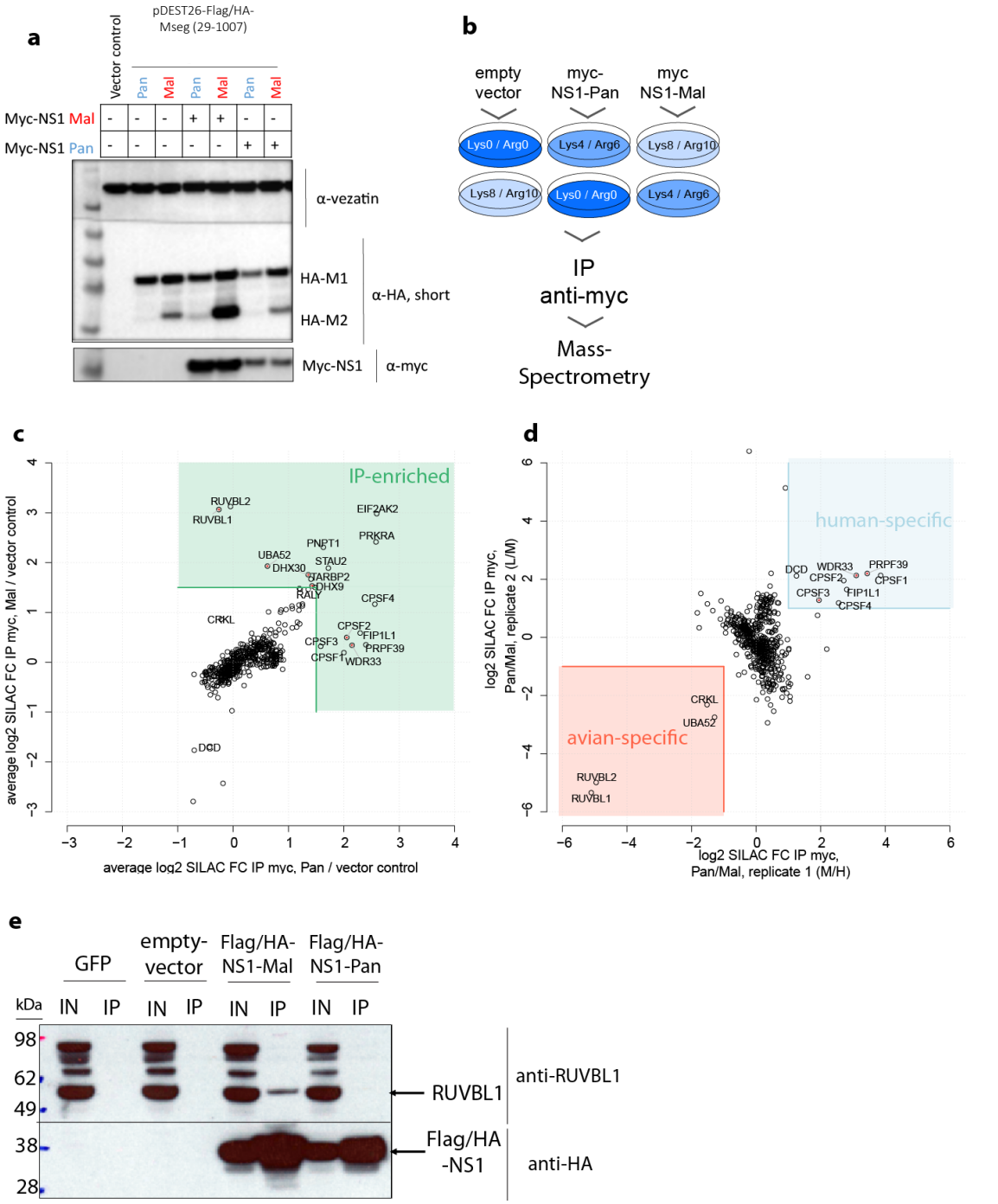


Figure 21 NS1 is a trans-acting factor controlling M segment RNA splicing.

(a) *myc-NS1 of Mal and Pan were transfected with M segment reporter constructs into A549 cells. Cells were harvested 24 hours post transfection. The levels of the indicated proteins were assessed by immunoblotting.* ***(b)*** *SILAC L/M/H cells were transfected with myc-NS1 expression constructs. 24 hours post transfection, cells were harvested, immunoprecipitations were performed and samples were analyzed by mass-spectrometry.* ***(c)*** *The axes in the scatterplot give the enrichment of proteins comparing enrichment with the NS1 variants to empty vector control purifications. The mean of n=2 replicates is given.* ***(d)*** *The SILAC comparison of proteins enriched with both NS1 variants for both replicates.* ***(e)*** *The indicated plasmids were transfected into HEK293T cells and anti-HA immunoprecipitations were performed. Input and IP (after enrichment) fractions were analyzed for RUVBL1 and Flag/HA-NS1 levels via immunoblotting. The experiment in a was performed by Martha Hergeselle. In addition Katrin Eichelbaum contributed to the raw-data of c and d.*

6. DISCUSSION

Advances in high-throughput sequencing have provided insights into the extraordinary diversity of viruses and their genomic determinants of host adaptation. However, the mechanism how these adaptive mutations enable replication in a given host is less well understood. Thus, systematic studies are required to understand how these related viruses reach different infection outcomes. Here we present the first systematic assessment of protein synthesis at global scale upon IAV infection. We compared infection of an avian-adapted virus to an infection with a human adapted virus in human cells to find that differential synthesis of individual viral proteins can underpin host adaptation. Below I discuss, the limitations and benefits of the chosen proteomic approach. Further, I discuss the findings in the context of IAV infection and host adaptation.

6.1. Limitations of pAHA labeling

Despite the advantages of pAHA-SILAC for studying proteome dynamics, it is important to keep its limitations in mind. First, the synthesis profiles that were recorded are not comprehensive. This means that some proteins may have escaped identification and quantification, because they are lowly abundant and/or lack tryptic peptides. Second, AHA was described to have some effect on protein abundance and on global translational activity when administered to cells for prolonged time intervals (24 h, (Bagert et al., 2014). For this reason, we only used relatively short pulse periods (4 h). However, we cannot exclude that AHA had some effect on protein homeostasis.

Third, as for all enrichment-based approaches it is critical to consider that some proteins may bind to the alkyne beads without having incorporated AHA. This may in particular be the case for proteins having many free thiol groups, such as proteins containing many cysteines. In that case, the alkyne beads may bind to thiol groups by a “thiol-yne” reaction (Lowe, 2014). However, the relatively good correlation between pAHA-SILAC and RNA-level data indicate that the extent of this problem is low in our data, especially for proteins that are induced upon infection (Figure 12). In any case, for future studies it may be advisable to introduce some changes into the enrichment protocol. For example, reduction and alkylation of cysteine residues could be performed prior to the enrichment step. Also, one could use another biorthogonal amino acid such as Homopropargylglycine (HPG) instead of or in addition to AHA (Calve et al., 2016). Similar to AHA, HPG is analogous to methionine but contains an alkyne group instead of an azide.

Lastly, during the process of analyzing the data, we observed that modified peptides from viral proteins were sometimes mistakenly interpreted by MaxQuant as derived from host proteins. This biased the quantification of the corresponding host protein. There are two reasons for this: i) Using pAHA, we compared two proteomes (viral and host) with vastly different abundances. Viral proteins were on average about ~200 times more abundant than cellular proteins (Figure 11). ii) In shotgun proteomics, peptides occur not only in their unmodified state but also as modified versions that are typically less abundant (e.g. peptides that have deamidated glutamine and asparagine residues) (Nielsen et al., 2006). When these modifications are not considered during a database search the search engine may misinterpret them as peptides with a

different sequence (Bogdanow et al., 2016). Thus, it is important to make sure that the search space is comprehensive when using pAHA-labeling and infection. In our case, we could almost completely eliminate false quantifications due to misassigned modified peptides by considering deamidation as a variable modification during the database search.

6.2. The dynamic host proteome

In the presented thesis it was assessed how a permissive infection of a human-adapted IAV isolate compares to non-permissive infection with an avian-adapted IAV isolate. The human virus efficiently grows in human epithelial cells but the avian virus reaches only very low titers (Figure 6 and Figure 18). This restriction must be intracellularly as both viruses enter cells efficiently (Figure 15) and start their gene expression program (Figure 13). Thus, we attempted to measure protein synthesis globally for many host and viral proteins comparing both strains. This approach is conceptually very similar to labeling approaches with ^{35}S methionine. Such approaches have been performed with Influenza A viruses during the 80's (Inglis, 1982; Inglis and Brown, 1984; Israel, 1980; Shapiro et al., 1987). Of note, these older approaches are necessarily limited in depth, as they are exclusively detecting only a few viral proteins due to the IAV induced host shutoff. Using pulse labeling with AHA, we could now identify synthesis of thousands of host proteins during IAV infection. In terms of protein mass, the total amount of proteins (host and viral) produced by the cell drops to ~1/4 during later stages (Figure 10), which may reflect a global stress

response that is induced by infection. At the molecular level, such global down-regulation in protein synthesis output can be caused by phosphorylation of EIF2α (eukaryotic translation initiation factor 2 subunit alpha) by cellular kinases that sense viral infection such as PKR (Stern-Ginossar et al., 2019). Although IAV counteracts the PKR-mediated phosphorylation of EIF2α (Tan and Katze, 1998), our results may indicate that the cellular stress response is still functional.

Most of the host proteins are produced much weaker during infection than during the uninfected state (Figure 11). This host shutoff has been attributed to the endonuclease activity of the PA-X (a cytosolic isoform of PA (Jagger et al., 2012)) and PA (component of the viral polymerase that resides in the nucleus) proteins that downregulate host mRNA half-lives. Concomitantly, viral mRNAs levels increase so that mRNA pool is predominantly composed of viral mRNAs (Bercovich-Kinori et al., 2016). We observed that mRNA and protein synthesis level data correlate well (Figure 12), which is line with the view that mRNA levels are determining the IAV-induced host shutoff (Bercovich-Kinori et al., 2016; Jagger et al., 2012). Interestingly, the correlation of mRNA with synthesized proteins was weaker for the human-adapted than for the avian-adapted strain. This may indicate an additional layer of post-transcriptional control that is exerted during infection with human-adapted IAV. For example, the interaction of the CPSF complex with the human-adapted NS1 (Figure 21) is likely to interfere with polyadenylation of host pre-mRNAs (Nemeroff et al., 1998).

Interestingly, some of the host proteins were not subject to shutoff in host protein synthesis. This included proteins related to translation, mitochondrial proteins and interferon-related proteins (Figure 11). This was due to upregulation of the

corresponding mRNA levels (Figure 12). It appears that the primary battleground that determines whether defense proteins of the host (i.e. interferon related proteins that counteract viral infection) or viral proteins are produced manifests at the transcriptional rather than translational level.

In the uninfected state, cellular proteins that are highly abundant are also synthesized more efficiently than low abundant proteins (Schwanhausser et al., 2011). Also, highly expressed genes tend to have more optimal codon content across species, suggesting that they have evolved to be highly efficiently translated (Man and Pilpel, 2007). However, the data presented indicates that the highly abundant viral proteins are not more efficiently synthesized than host proteins (Figure 12c), which is consistent with previous reports (Bercovich-Kinori et al., 2016). In consistence with this idea, computational analyses on many IAV sequences show that the codon usage of IAV genes is not optimized to reflect the codon usage of the host (Zhou et al., 2005). However, these global analyses contradict early studies based on reporter systems that identified sequences in 5'UTRs (5' untranslated regions) of viral mRNAs that were proposed to mediate selective translation of viral over host mRNAs (Garfinkel and Katze, 1992, 1993).

6.3. Dysregulated synthesis of viral proteins

We observed that protein production of IAV is severely dysregulated during a non-permissive infection (Figure 14 and Figure 20). This is in contrast to the synthesis of host proteins that was overall similar between both strains (Figure 11). This suggests that differences in viral rather than host proteins determines the outcome of the infection. Compared to the human-adapted virus the avian-adapted virus produced at the beginning all viral proteins stronger. Then almost all proteins are produced less and less efficiently by the avian virus as the infection progresses. The global tendency towards inefficient production of viral proteins as the infection progresses reflects probably the impairment of the avian-adapted IAV polymerase in human cells (Gabriel et al., 2005; Long et al., 2016; Long et al., 2019). However, we observed that some proteins do not follow that general trend or are disproportionately affected. For example, the HA protein is stronger produced by the avian-adapted virus. This did not correlate with induced activation of MAPK-responsive genes (Figure 15) during Mal-infection which would be expected from increased HA membrane accumulation (Marjuki et al., 2006). Of note, this does not rule out that differences in HA protein levels play a role for crossing the species barrier. In particular, a balance between HA and NA is required for a permissive infection due to their complementary activities during budding (Wagner et al., 2002). Our observations that the NA protein was produced inefficiently in A549 cells by the avian-adapted IAV (Figure 14 and Figure 20) may indicate that progeny virions are aggregated, as previously observed for temperature sensitive NA mutants (Palese et al., 1974).

The strongest difference that we observed for individual proteins/mRNAs comparing the avian-adapted with the human-adapted virus was for the M1 protein. M1 is among the most conserved proteins/mRNAs at both the nucleotide and protein level, when comparing the human-adapted and the avian-adapted strain (Figure 22). M1 controls

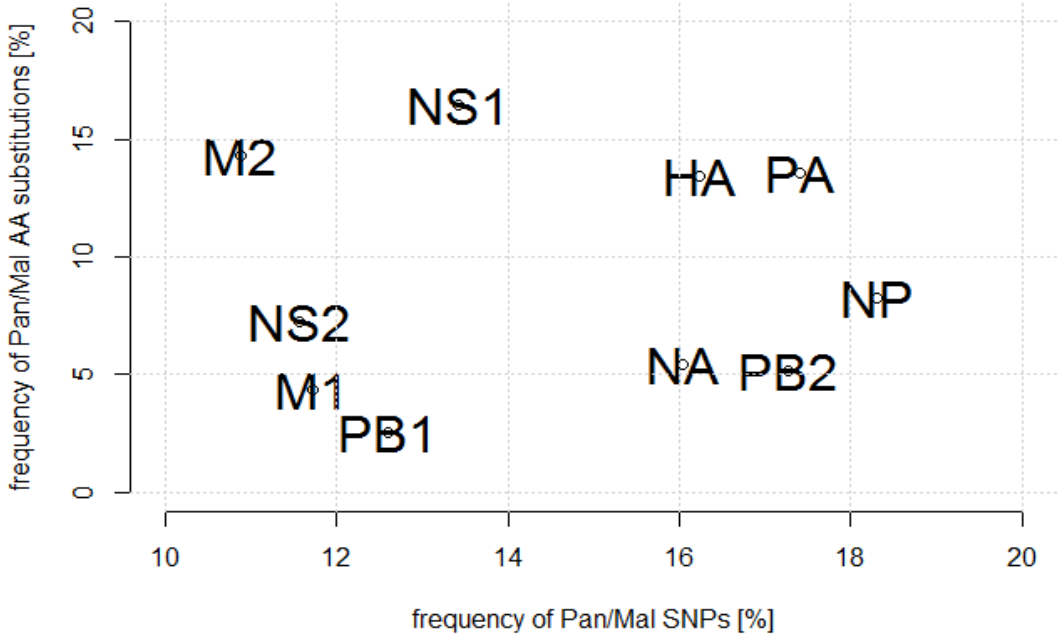


Figure 22 Differences of IAV encoded proteins at the nucleotide level (x-axis) and the protein level (y-axis), when comparing Mal and Pan strain.

the export of vRNPs from the nucleus to the cytosol (Bui et al., 2000; Martin and Helenius, 1991a). Consistently, during Mal infection the export of NP to the cytosol was impaired (Figure 15).

This phenotype (low M1, impaired NP export) is reminiscent to observations made in earlier studies under different non-permissive experimental infection settings. For example, with the non-neuroadapted strain A/WS/33 in mouse brain (Schlesinger et al., 1989). Also, with strain WSN when Ehrlich tumor cells were infected (Bukrinskaya

et al., 1981). In agreement with the data presented in this thesis, some researchers observed inefficient production of M1 when the FPV strain of avian influenza was used to infect mouse L (Inglis and Brown, 1984) or Vero cells (Lau and Scholtissek, 1995). Collectively, these and my observations pinpoint towards functional correlation of non-permissive infections with impaired production of M1 protein. Importantly, this correlation appears to be conserved across different non-permissive cellular experimental settings.

6.4. M segment RNA splicing

This thesis presents evidence that M segment RNA splicing is markedly different between non-permissive and permissive IAV infections (Figure 16). The splicing efficiency is controlled by *cis*-acting signals (that is, the 3' splice site, Figure 17, Figure 18) and *trans*-acting factors (i.e. NS1, Figure 21). IAV inherently depends on the host cell for splicing. This is in contrast to other RNA processing steps that are catalyzed by IAV proteins, such as transcription (Eisfeld et al., 2015), polyadenylation (Luo et al., 1991) and capping (Plotch et al., 1981). Thus, species-specific differences in host-specific *trans*-acting factors are likely to yield differences in alternative splicing of IAV RNAs.

This thesis proposes that the avian strain is adapted to its natural host cells in such a way that they permit high level of M1 protein production (Figure 20). In the wrong host, M1 is insufficiently produced due to – at least partially – high degree of splicing of the primary M transcript. This implies a suboptimal adaptation to the human splicing environment of avian-adapted strains. Indeed, a large-scale RNA-sequencing analysis of many vertebrate tissues identified diverging regulation of pre-mRNA splicing between humans and avian species (e.g. chicken) (Barbosa-Morais et al., 2012). In fact, splicing profiles diverge more between humans and chicken than gene expression levels (Barbosa-Morais et al., 2012). An example that is related to IAV host range is splicing of ANP32A, which is differently spliced in human and avian species (Baker et al., 2018). This creates avian-specific ANP32A isoforms that enhance IAV polymerase activity in avian cells (Long et al., 2016).

Further, this thesis shows that strain-specific M segment RNA splicing differences are controlled by the 3' splice site region (Figure 17). The importance of this *cis*-element is underscored by several points: i) It is the common acceptor site for all alternatively spliced transcripts of the M segment (Figure 16a). ii) It is of RNA secondary structure (Figure 18 and (Moss et al., 2012)). iii) It was shown to bind to the ASF/SF2 (SRSF1) splicing regulatory protein (Shih and Krug, 1996). iv) The sequence encodes M1 and M2 proteins that have distinct reading frames, which implies strong constraint at the nucleotide level. v) The sequence contains splicing regulatory signals such as the branch point and the polypyrimidine tract (Moss et al., 2012).

We found that the 3' splice site contains different RNA secondary structure elements that are predicted based on the sequence conservation across hundreds of avian- and human-adapted strains (Figure 18). Consistently, sequences containing the structured region around the 3' splice site migrated differently in native but not denaturing gels (Figure 18). The hairpin-like sequence that we identified in avian sequences is very similar to a structure proposed by Moss et al. for four avian sequences (Moss et al., 2012). Moss et al. further proposed that this sequence may exist in an equilibrium between hairpin and pseudoknot. Further, it was suggested that this equilibrium may change depending on the frequency of GC base pairings at the 3' splice site region, with avian IAV strains having higher frequency of GC base pairs. In consistence with this hypothesis, this thesis proposes that the region around the 3' splice site forms distinct structural elements depending on the host species of the IAV strain. However, the precise architectures of avian and human IAV 3' splice sites vary between the host species in our analysis. The reason for this discrepancy between this study and the

analysis by Moss et al. could be the different methodological approaches. While the study of Moss et al., investigated isolated fragments of the M segment RNA using thermodynamic stability modelling, our approach performed predictions based on the entire M segment RNA in their evolutionary context.

It will be interesting to directly link the secondary structure of the avian M segment at the 3' splice site to increases in splicing efficiency. One may speculate that this structural element exposes the splice site, which facilitates binding of splicing components. The picture is likely more complicated as it is known that structured splice acceptor sites can promote splicing (Coleman and Roesser, 1998), but also where a structured 3' splice site inhibits splicing (Watakabe et al., 1989).

Importantly, M-segment splicing does not only depend on *cis*-regulatory elements but also on *trans*-acting factors, such as NS1, RdRp, NS1-BP or HNRNPK (Mor et al., 2016; Robb and Fodor, 2012; Shih and Krug, 1996; Shih et al., 1995; Thompson et al., 2018). In fact, while M1 protein production was clearly impaired with our mutant strain (Figure 19), the wild type avian strain produced even less (Figure 14). It is therefore important to interpret the findings in the broader context of viral and host factors that jointly regulate M segment RNA splicing. In fact, it is shown that NS1 is one of these *trans*-acting factors that can promote M segment RNA splicing (Figure 21). Additional host factors that could be involved are RUVBL1/2 that co-precipitated with the Mal but not the Pan NS1. Considering that RUVBL1/2 are restriction factors of many viral infections (Yasunaga et al., 2014), are known to bind viral proteins (Morwitzer et al., 2019) and are involved in regulation of splicing (Cloutier et al., 2017), it may be worth considering that the NS1-RUVBL1/2 axis acts in *trans* to modulate M

segment RNA splicing. In support of this idea, experiments by Madlen Luckner (Herrmann lab, Humboldt University, personal communication) indicate that siRNA mediated knockdown of RUVBL1 or RUVBL2 enhances titers of Mal-virus infected A549 cells ~10 fold.

6.5. Implications for host range

The infection model that was used here is advantageous as it allows experiments under well-controlled conditions. Nevertheless, it is important to state that the strains we used do not represent the full diversity of IAVs in humans or birds. However, several lines of evidence suggest that the strain-specific control of M segment RNA splicing by the 3' splice site is conserved and important for host range.

First, the computational analysis of avian versus human M segment RNA secondary structures is based on hundreds of avian and human sequences (Figure 18). RNA secondary structures are only identified when they are widely conserved across these sequences. Second, phylogenetic trees of M segment show clear host-specific clustering (Furuse et al., 2009). Third, the analysis of an early avian virus isolate (FPV) from the 1930's shows that this M segment is very efficiently spliced. Importantly, mutation of the 3' splice towards the human sequence completely shifted the splicing pattern towards very inefficient splicing. Fourth, the analysis with the p1918 virus that caused the "Spanish flu" outbreak. This virus spread globally and contributed segments to all pandemic outbreaks that were recorded since (Morens et al., 2009). Consistent with previous reports (Backstrom Winquist et al., 2012), we show that the M segment of this virus is spliced poorly (Figure 18). Interestingly, the M segment of this virus is in some regions similar to avian-adapted strains but shows signatures of mammalian adaptation especially at the 3' splice site (Reid et al., 2002). The M gene of the Pan virus that was assessed in this study was contributed by the p1918 virus. Consistently, both p1918 and Pan strain splice the M segment poorly (Figure 18).

Fifth, avian and human viruses differ by five nucleotides at the 3' splice site, all of which are retained with the Pan and Mal strains employed here (Figure 23). Thus, Pan and Mal strains are by and large representative for avian and human strains at the 3'splice site. Sixth, mutating eight nucleotides at the 3' splice site of the Pan strain to the corresponding sequence of the Mal virus impaired growth of the Pan strain about 10-fold (Figure 19). In addition, recent results from the Steel lab show that avian M segments restrict growth and transmission of mammalian-adapted IAV strains in a guinea pig model system (John Steel, Emory University, personal communication). Collectively, these points strongly suggest that the results are relevant outside the specific experimental model system employed here and that they are important for host range.

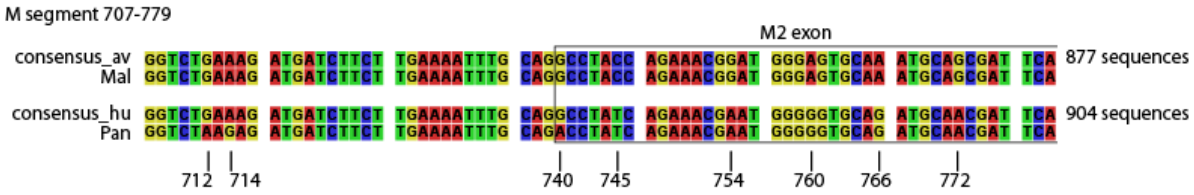


Figure 23 Alignment of consensus sequences of avian ("consensus_av") and human ("consensus_hu") M segment 3' splice site regions with Pan and Mal sequences.

A hypothetical model on the role of M segment RNA splicing is sketched in Figure 24. Avian-adapted IAV M segment pre-mRNAs contain a hairpin-like RNA secondary structure at the 3' splice site while human-adapted sequences contain a different RNA secondary structure. Splicing of the human-IAV adapted M segment RNA is inefficient leading to high levels of M1 mRNA and protein. High M1 protein level may facilitate export of viral RNPs to the cytoplasm. In contrast, different M segment RNA secondary structures predominate with avian IAVs and splicing is very efficient. This leads to

underproduction of M1 mRNA and protein, which may impact the proper transport vRNPs to the cytosol.

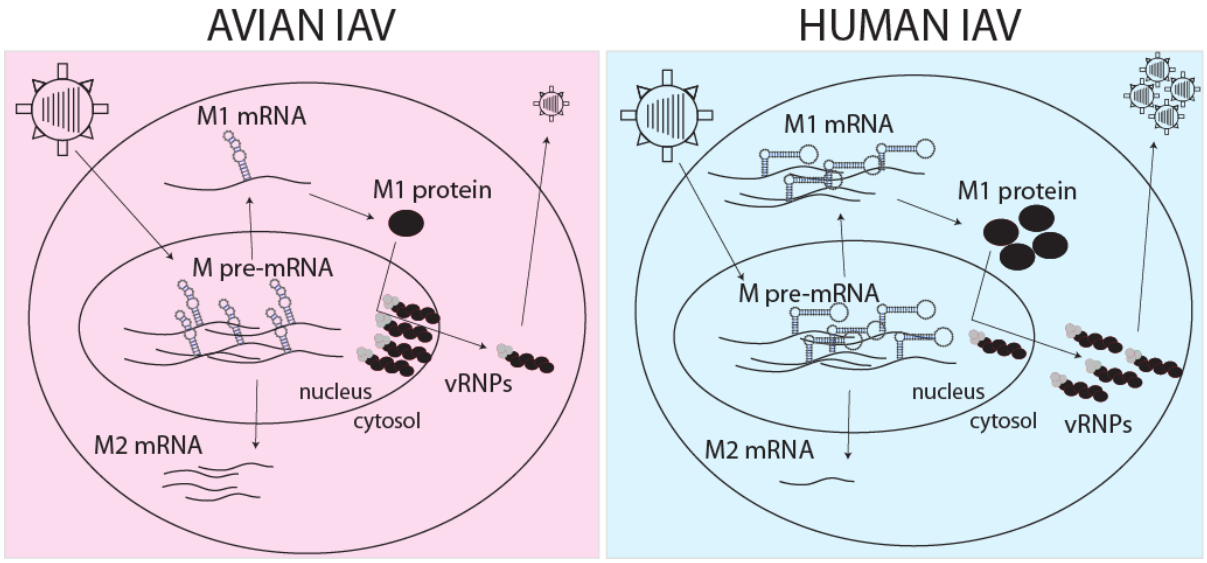


Figure 24 Hypothetical model for the role of M segment splicing for IAV host range.

Different RNA secondary structure elements predominate at the 3' splice site in avian and human IAV M pre-mRNAs. The structural element in avian IAVs may facilitate splicing, which leads to the underproduction of M1 mRNA and protein in human cells infected with avian-adapted IAVs. The poor availability of the M1 protein may contribute to an impaired nuclear export of vRNPs.

7. OUTLOOK

7.1. Potential impact of the work

So far, the M segment plays a rather neglected role in the focus of researchers investigating IAV host adaptation. Hopefully, the presented data will provide incentive into further researching adaptive mechanisms of IAVs with respect to splicing of the M segment. This will not be a trivial task due to the diversity of IAVs and interconnected mechanisms of host adaptation. In any case, it may be advisable to keep modulation of M segment RNA splicing in mind, when screening for adaptive mutations harbored by IAV isolates. However, these adaptive mutations can but most not arise in the 3' splice site region. Adaptive mutations modulating M segment RNA splicing could be relatively difficult to spot due to their involvement in RNA secondary structures.

This work shows that differences in protein expression kinetics can be important for host range and that these differences underpin host adaptation. We expect that the principle tools to investigate this problem (pSILAC, pAHA, AP-MS) will be more broadly used by virologists in the future also for other infection systems.

pAHA allowed us to enrich viral proteins over host proteins under non-ectopic environments. This can be a clear advantage for researchers investigating virus host protein interactions on a global scale over more traditional AP-MS approaches. For example, one could combine pAHA-labeling and cross-linking mass spectrometry. This would allow us to assess the interactions of host proteins with all viral proteins under non-ectopic environments in single-shot experiments.

7.2. Open questions

This work has generated numerous questions that may be addressed in the future. Some of them are listed below:

1) Are some of the proteins least subject of host shutoff required by the virus and induced for this reason or does their induction reflect a global cellular defense response to the pathogen?
2) How do cellular and host factors functionally interact to produce the host shutoff? That means, what is the contribution of the cellular defense and that of IAV proteins? Also, which viral proteins (i.e. PA-X, PA, NS1) are most critical in promoting host shutoff?
3) Is the dysregulation of NS-segment RNA splicing also important for host range, as suggested by previous studies (Huang et al., 2017)? The question is interesting as NS1 is a positive splicing regulator that acts on M segment RNA splicing and NS2 can promote export of vRNPs to the cytosol.
4) Is the NP-export defect observed with the Mal strain a direct consequence of low M1 levels or is it a by-product of a different block to avian-adapted IAV?
5) What is the precise role of host and viral *trans*-acting host factors in modulating splicing efficiency of the M segment and what are differences in the mechanisms by which birds and humans splice pre-mRNAs? Specifically is the NS1-RUVBL1/2 axis involved in controlling the levels of splicing?
6) Would an avian virus with a human splice site region show an increased ability to produce novel progeny?

7) The M segment has adapted to humans twice in the pandemic history of IAV. For the first time in 1918 with the “Spanish flu” (Worobey et al., 2014a) and for the second time in the genesis of the 2009 swine origin Influenza pandemic (Smith et al., 2009). The Pan strain that we investigated gained its M segment from the 1918 pandemic virus. It will be interesting to also investigate adaptive mechanism with regard to M segment splicing in the context of the 2009 swine origin Influenza as well.

Most importantly, an increased understanding of M segment RNA splicing and its host and viral determinants will hopefully increase our abilities for risk assessment of potential pandemic events of IAV.

8. SUPPLEMENTARY INFORMATION

8.1. Recurring abbreviations

Abbreviation	Meaning
ABC	ammonium bicarbonate
AHA	L-azidohomoalanine
AIV	avian influenza virus
ANP32A	acidic nuclear phosphoprotein 32 family member A
AP-MS	affinity purification mass-spectrometry
Arg	arginine
ATCC	American type culture collection
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CPSF	cleavage and polyadenylation specificity factor
Ct	cycle threshold
ddH_2O	double distilled water
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FFU	fluorescence forming unit
FITC	fluorescein isothiocyanate
FPV	fowl plague virus
FSC	forward light scatter
fw	forward
GO	gene ontology
HA	hemagglutinin
HCV	hepatitis C virus
HIV	human immunodeficiency virus

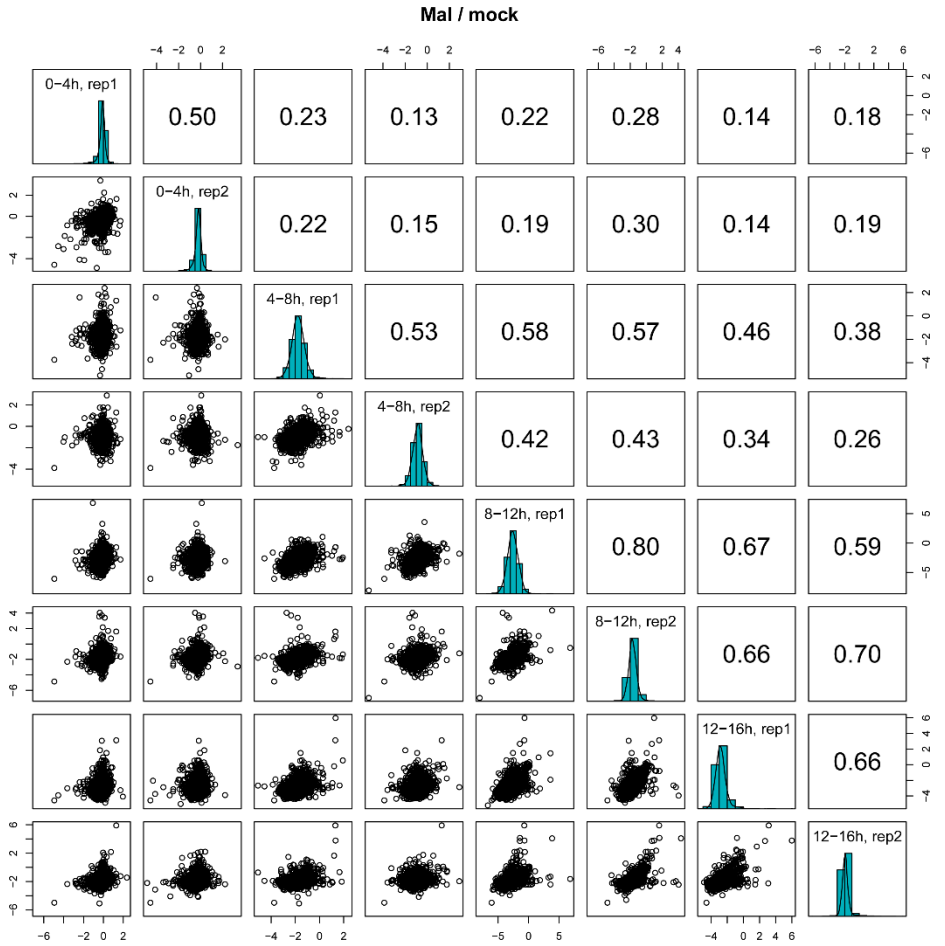
HPAI	high pathogenic avian influenza
HPG	Homopropargylglycine
hpi	hours post infection
HNRNPK	Heterogeneous Nuclear Ribonucleoprotein K
IAA	iodoacetamide
IAV	Influenza A virus
iBAQ	intensity based absolute quantification
IFN	interferon
ISG	interferon stimulated gene
LB	lysogeny broth
LPAI	low pathogenic avian influenza
Lys	lysine
m/z	mass divided by charge
M1	matrix-1 protein
M2	matrix-2 protein
M4	matrix-4 protein
Mal	A/Mallard/439/2004
MAPK	mitogen activated protein kinase
MEM	minimum essential medium
Met	methionine
MOI	multiplicity of infection
mRNA	messenger ribonucleic acid
MS	mass-spectrometry
MxA	Mx dynamin like GTPase 1
NA	neuraminidase
NCBI	national center for biotechnology information
NP	nucleoprotein
NS1	non-structural protein 1
NS1-BP	NS1 binding protein
NS2/ NEP	non-structural protein 2/ nuclear export protein
nt	nucleotide

p1918	A/Brevig Mission/1/1918
PA	polymerase acidic protein
Pan	A/Panama/2007/1999
PB1	polymerase basic protein 1
PB2	polymerase basic protein 2
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFU	plaque forming unit
pH	potential of hydrogen
PKR	protein kinase R
pol II	RNA polymerase II
pSILAC	pulsed stable isotope labeling of amino acids in cell culture
qRT-PCR	quantitative real-time polymerase chain reaction
RdRp	RNA dependent RNA polymerase
rev	reverse
RNA	ribonucleic acid
RPKM	reads per kilobase per million reads
RUVBL1	RuvB Like AAA ATPase 1
RUVBL2	RuvB Like AAA ATPase 1
rRNA	ribosomal RNA
SILAC	stable isotope labeling of amino acids in cell culture
SRSF1	serine and arginine rich splicing factor 1
SSC	sideward light scatter
SV40	simian virus 40
TFA	trifluoracetic buffer
tRNA	transfer RNA
vRNA	viral RNA
vRNPs	viral ribonucleoproteins
w/	with
w/o	without
WHO	world health organization

8.2. Raw data access

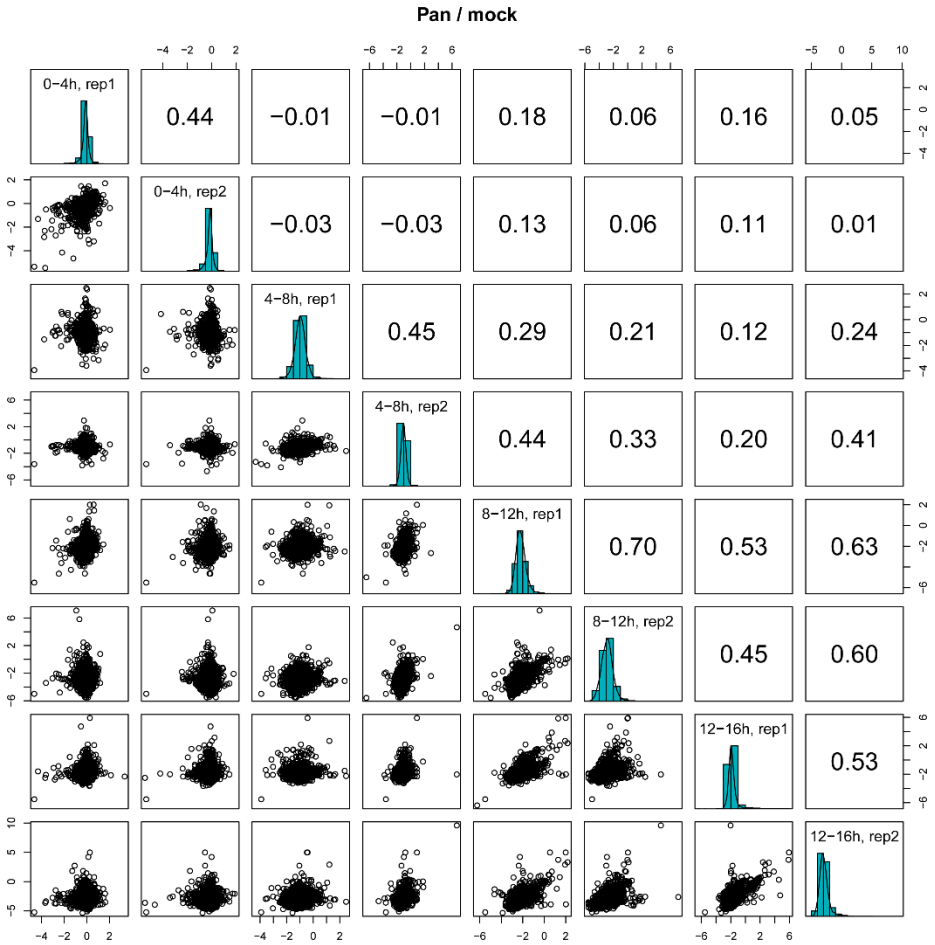
Proteomic data relating to the pAHA-SILAC experiment was uploaded to ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD011321 (username: reviewer41195@ebi.ac.uk and password: WjfJjeMY). RNA-sequencing data is available at https://dataview.ncbi.nlm.nih.gov/object/PRJNA495615?reviewer=tl7nog5od602n7c69e1neapet and will become available publicly via the BioProject identifier PRJNA495615 via the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA495615.

8.3. Supplementary Figures 1-7



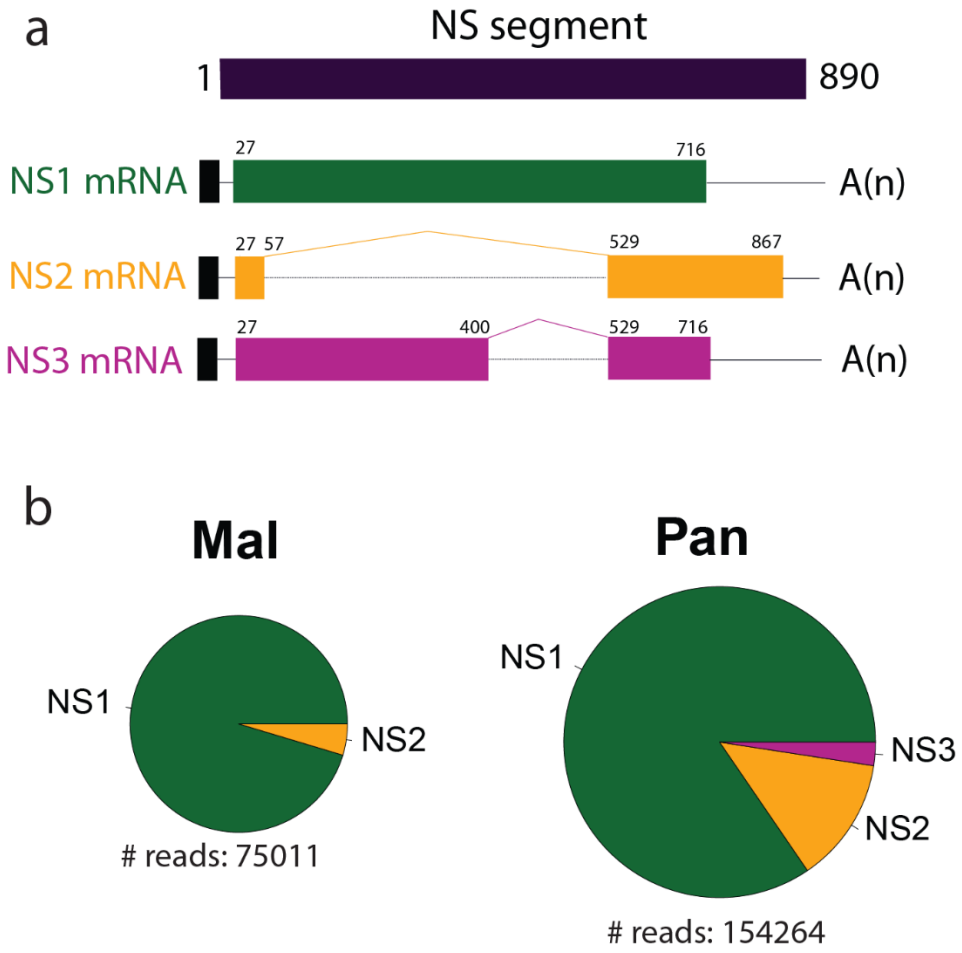
Supplementary Figure 1 Spearman's correlation coefficients and scatter plots of biological (label-swap) replicates for SILAC ratios of the Mal/mock infection conditions.

In the first replicate SILAC H labelled cells were infected with the Mal strain, SILAC M labelled cells infected with the Pan strain, and SILAC L cells were mock infected. In the second replicate, SILAC M labelled cells were infected with the Mal strain, SILAC H labelled cells were infected with the Pan strain and SILAC L cells were mock infected.



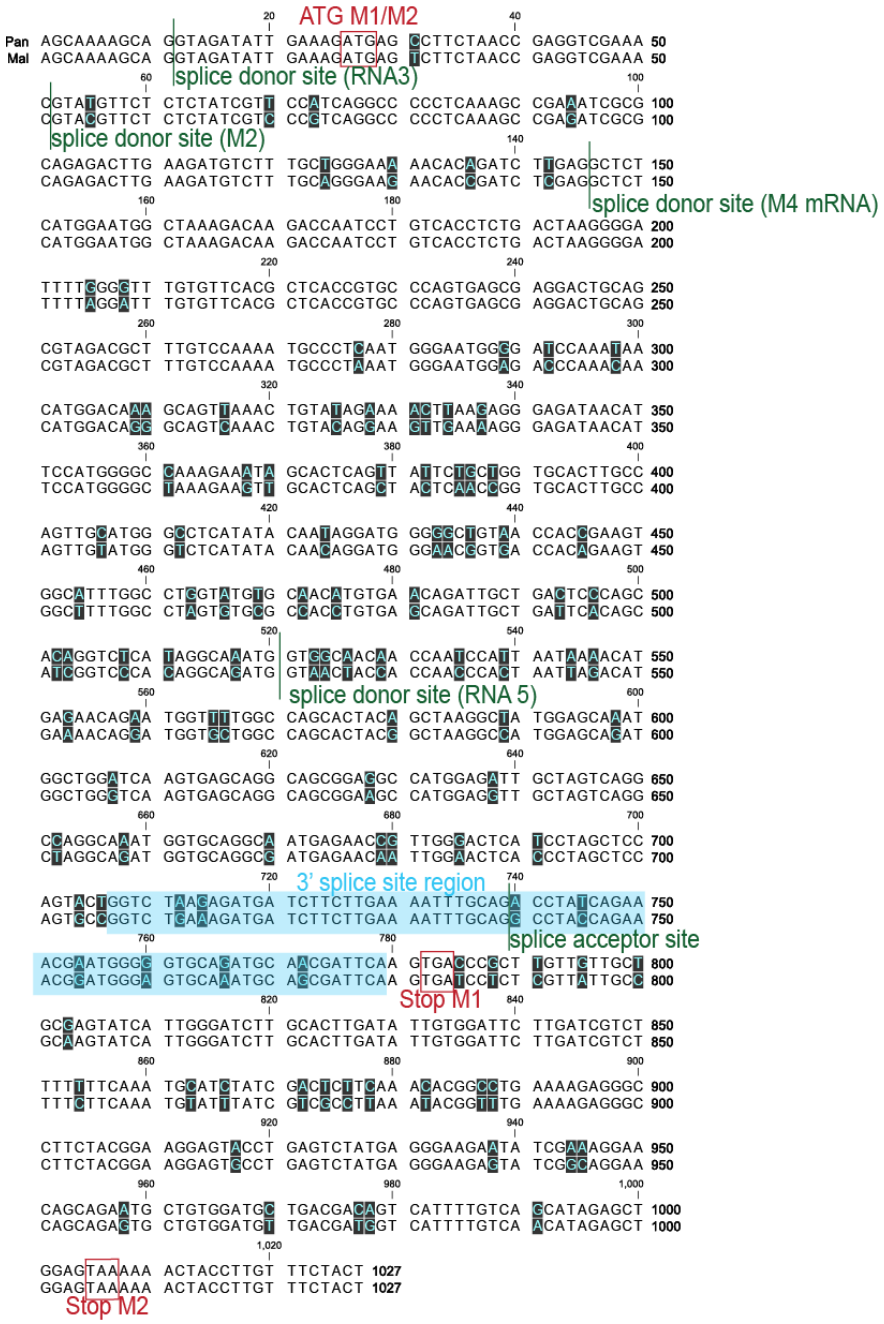
Supplementary Figure 2 Spearman's correlation coefficients and scatter plots of biological (label-swap) replicates for SILAC ratios of the Pan/mock infection conditions.

In the first replicate SILAC H labelled cells were infected with the Mal strain, SILAC M labelled cells infected with the Pan strain, and SILAC L cells were mock infected. In the second replicate, SILAC M labelled cells were infected with the Mal strain, SILAC H labelled cells were infected with the Pan strain and SILAC L cells were mock infected.



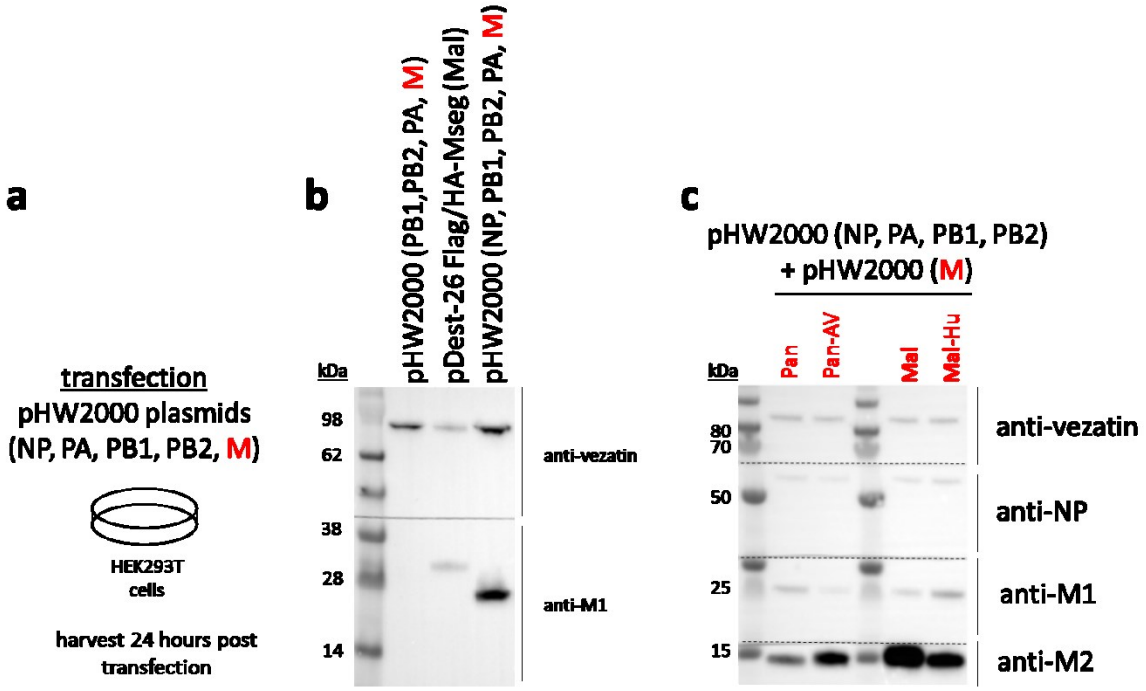
Supplementary Figure 3 Comparative Assessment of NS segment RNA splicing.

(a) *Schematic depiction of NS gene architecture. The NS segment is the second-smallest segment of IAV and encodes three major transcripts: The collinear transcript encoding NS1 (green), and two alternatively spliced transcript (NS2 mRNA, yellow) that encodes for NS2 and NS3 mRNA (purple). The ORFs are indicated with colored boxes. In addition, the positions of the ATG codon, of splice donor and acceptor sites as well as stop codons are indicated.* ***(b)*** *Relative quantification of the different isoforms based on splice junction reads from RNAseq data for both strains at 8 hpi. The area of the pies reflects the absolute number of splice junction reads. The mean of two replicates is depicted The analysis quantifying splice junction reads was performed by Xi Wang (MDC Berlin)*



Supplementary Figure 4 M segment nucleotide sequence alignment of Pan (top sequence) and Mal (bottom) strain.

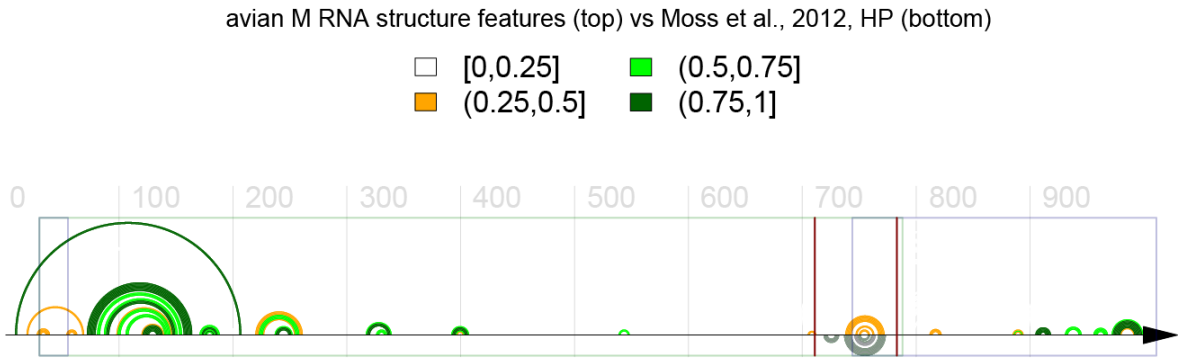
Splice donor and acceptor sites are indicated with green bars. ATG and stop codons of M1 and M2 ORFs are highlighted red. The 3' splice site region responsible for strain-specific splicing is shaded cyan. Nucleotide positions with different sequence between Pan and Mal strain are depicted on black background.



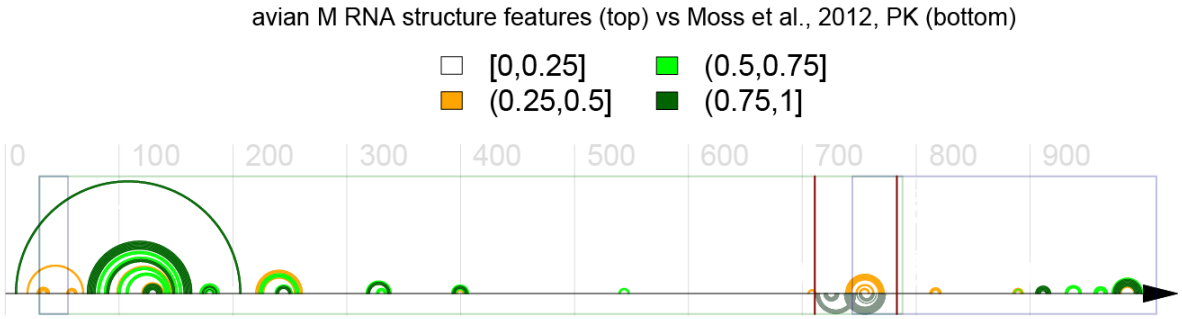
Supplementary Figure 5 The 3' splice site region controls M segment RNA splicing in an RdRp reporter system.

(a) *pHW2000 plasmids encoding the NP, PA, PB1, PB2 (of strain WSN) and M segments of Pan and Mal and 3' splice site chimeric constructs was transfected to HEK293T cells. In this system, the accumulation of M1 depends on the activity of the RNA-dependent RNA polymerase.* ***(b)*** *The plasmids as indicated were transfected into HEK293T cells. In the absence of the pHW2000 NP plasmid (lane 1), M1 did not accumulate. In the presence of NP, M1 did accumulate (lane 3). This indicates that M1 accumulation depends on the activity of the RdRp. Lysates of cells transfected with a pDEST26-Flag/HA-Mseg(29-1007) plasmid (lane 2) serve as positive control for M1 immunoblots.* ***(c)*** *pHW2000-plasmids expressing NP and polymerase subunits were transfected into HEK293T cells together with the indicated pHW2000 M segment plasmids of strain Pan, Mal or chimeric constructs. Pan-AV denotes the Panama M segment with avian IAV 3' splice site and Mal-Hu denotes the Mal M segment with human IAV 3' splice site. Note that the antibody we used for M2 does not allow to draw conclusions about the relative amount of M2 comparing Pan and Mal strain. This is due to this antibody being raised against an internal region of the M2 protein that is different between Pan and Mal strain. However, this antibody allows comparisons between the wild type and mutant M2 variants.*

a

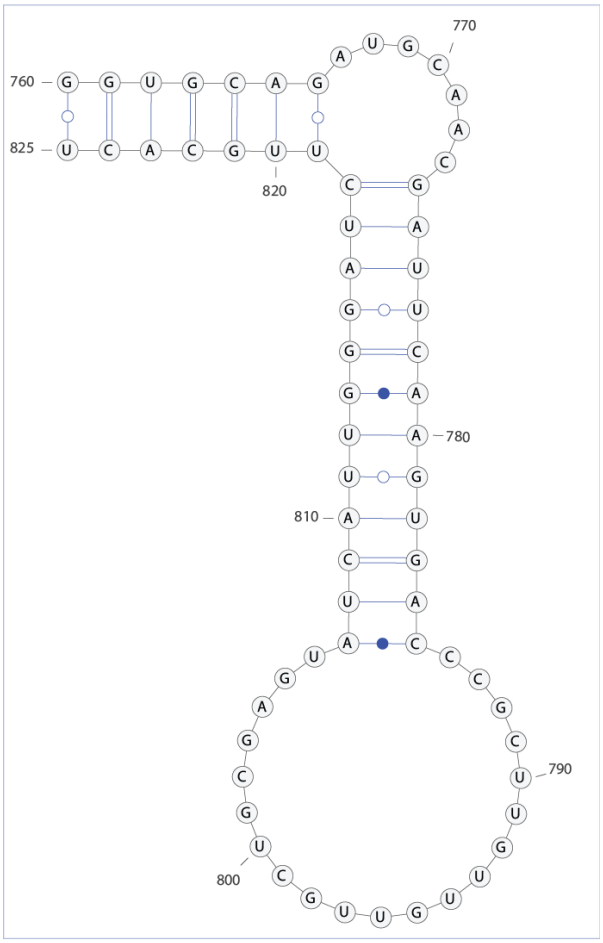


b



Supplementary Figure 6 RNA structure prediction comparison to published data.

*Agreement between the RNA secondary structure predicted by RNA-Decoder for the M segment RNA for the avian-adapted influenza strains and the RNA structure features predicted by Moss et al., PloS One, 2012. Agreement with the hairpin (HP) in **(a)** and with the pseudoknot (PK) in **(b)**. Color-coding of base-pairs according to the corresponding, estimated base-pairing probabilities, see also legend of Figure 18. The single exon containing the contiguous open-reading frame (ORF) of the M1 splice variants is indicated by a green box, the two exons corresponding to the M2 splice variant are shown as two blue boxes. Note that there are two regions where the ORFs of the splice variants overlap. The first one corresponds to the first exon of the M2 splice variant where the two ORFs are in sync. The second one corresponds to the region of overlap between the 3' end of the long M1 ORF and the 5' start of the second exon of the M2 splice variant. In that region, the two ORFs are out of sync, implying a particularly strong constraint due to the two, intertwined amino-acid contexts. The two vertical red bars indicate the start and end position of the region that was used for chimeric constructs around the 3' splice site. This figure is courtesy of Irmtraud Meyer, MDC Berlin.*



Supplementary Figure 7 Visualization of the RNA secondary structure element predicted for human IAV isolates.

The region predicted to be of secondary structure with human IAV isolates (nucleotides 760-825) was plotted with the nucleotide sequence of the Pan strain.

8.4. Supplementary Table 1 and 2

Supplementary Tables 1-2 can be downloaded from:

https://www.biorxiv.org/content/10.1101/438176v1.supplementary-material

8.5. Supplementary Table 3

Supplementary Table 3 Proteins co-precipitating with NS1 of Mal or Pan strain and their respective log2 SILAC ratios in triple SILAC AP-MS experiments (related to Figure 21).

SILAC ratios in the affinity purification with Pan or Mal NS1 compared to the empty vector control (empty) on a green/white scale. SILAC ratios directly comparing Pan / Mal are scaled blue and red, respectively. Proteins are included that have an average log2 fold-change NS1 versus empty vector control larger than 1.The ratio for the protein DCD indicated strain-specific enrichment to NS1. However, it was not solidly identified as an interactor when comparing to the empty vector control. For this reason the quantification of DCD has been removed from this table.

Gene names	Pan/empty, rep 1	Pan/empty, rep 2	average Pan/empty	Mal/empty, rep 2	Mal/empty, rep 1	average Mal/empty	Pan/Mal, rep 1	Pan/Mal, rep 2	average Pan/Mal
CPSF1	3.70	2.30	3.00	0.19	0.39	0.29	3.86	2.13	2.99
PRPF39	3.49	3.70	3.60	1.31	-0.26	0.53	3.44	2.20	2.82
WDR33	4.25	2.19	3.22	0.12	0.90	0.51	3.10	2.13	2.62
NS1_Pan	10.04	2.35	6.19	0.56	0.39	0.47	3.28	1.89	2.59
CPSF2	3.33	2.82	3.08	0.71	0.79	0.75	2.71	1.95	2.33
FIP1L1	4.06	2.82	3.44	0.57	1.19	0.88	2.81	1.64	2.23
CPSF4	4.30	3.38	3.84	1.94	1.56	1.75	2.56	1.18	1.87
CPSF3	2.72	2.03	2.37	0.35	0.61	0.48	1.95	1.28	1.61
DHX9	2.97	1.46	2.22	2.68	1.83	2.26	1.02	-1.70	-0.34
TARBP2	2.72	1.50	2.11	2.31	2.70	2.50	-0.01	-0.88	-0.45
RALY	3.27	1.00	2.14	2.50	2.10	2.30	1.05	-2.04	-0.49
STAU2	3.77	1.39	2.58	2.92	2.75	2.83	0.81	-2.06	-0.63
EIF2AK2	4.40	3.37	3.89	5.00	3.95	4.47	0.36	-2.12	-0.88
PRKRA	5.12	2.62	3.87	4.44	2.80	3.62	0.45	-2.33	-0.94
DHX30	2.54	1.53	2.03	2.91	2.36	2.63	0.01	-2.14	-1.07
PNPT1	4.20	0.68	2.44	3.16	3.76	3.46	0.36	-2.94	-1.29
CRKL	-0.33	-0.33	-0.33	1.59	0.99	1.29	-1.52	-2.32	-1.92
UBA52	1.12	0.74	0.93	3.05	2.74	2.89	-1.29	-2.75	-2.02
RUVBL2	-0.10	-0.04	-0.07	4.80	4.58	4.69	-4.95	-4.99	-4.97
RUVBL1	-0.76	-0.02	-0.39	4.85	4.34	4.60	-5.10	-5.34	-5.22
NS1_Mal	0.42	-11.48	-5.53	3.29	6.99	5.14	-9.29	-7.46	-8.37

8.6. Publications

During PhD studies

Koshi Imami, Miha Milek, **Boris Bogdanow**, Tomoharu Yasuda, Nicolai Kastelic, Henrik Zauber, Yasushi Ishihama, Markus Landthaler, Matthias Selbach: *Phosphorylation of the Ribosomal Protein RPL12/uL11 Affects Translation during Mitosis*. Molecular Cell 09/2018; 72(1)., DOI:10.1016/j.molcel.2018.08.019

Anne Sadewasser, Katharina Paki, Katrin Eichelbaum, **Boris Bogdanow**, Sandra Saenger, Matthias Budt, Markus Lesch, Klaus-Peter Hinz, Andreas Herrmann, Thomas F Meyer, Alexander Karlas, Matthias Selbach, Thorsten Wolff: *Quantitative Proteomic Approach Identifies Vpr Binding Protein as Novel Host Factor Supporting Influenza A Virus Infections in Human Cells*. Molecular & Cellular Proteomics 03/2017; 16(5):mcp.M116.065904., DOI:10.1074/mcp.M116.065904

Boris Bogdanow*, Henrik Zauber*, Matthias Selbach: *Systematic Errors in Peptide and Protein Identification and Quantification by Modified Peptides*. Molecular & Cellular Proteomics 05/2016; 15(8):mcp.M115.055103., DOI:10.1074/mcp.M115.055103

*co-first authors

in revision

Bogdanow, Boris, Katrin Eichelbaum, Anne Sadewasser, Xi Wang, Immanuel Husic, Katharina Paki, Martha Hergeselle et al. *The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant*. bioRxiv (2018): 438176. (in revision at Nature Communications)

in preparation

Boris Bogdanow*, Max Schmidt*, Henry Weisbach, Iris Gruska, Barbara Vetter, Koshi Imami, Matthias Selbach, Chrisitan Hagemeier, Lüder Wiebusch: *The protein-protein Interactome of herpesviral kinases reveals cross-regulation with cyclin/cdk complexes*

*co-first authors

Prior to PhD studies

Martin Eifler, Ralf Uecker, Henry Weisbach, **Boris Bogdanow**, Ellen Richter, Lydia König, Barbara Vetter, Tihana Lenac-Rovis, Stipan Jonjic, Heidemarie Neitzel, Christian Hagemeier, Lüder Wiebusch: *PUL21a-Cyclin A2 Interaction is Required to Protect Human Cytomegalovirus-Infected Cells from the Deleterious Consequences of Mitotic Entry*. PLoS Pathogens 11/2014; 10(11):e1004514., DOI:10.1371/journal.ppat.1004514

Sameh Selim, Céline Roisin-Fichter, Jean-Baptiste Andry, **Boris Bogdanow**, Renée Sambou: *Real-time PCR to study the effect of timing and persistence of fungicide application and wheat varietal resistance on Mycosphaerella graminicola and its sterol 14 α -demethylation-inhibitor-resistant genotypes*. Pest Management Science 03/2014; 70(1)., DOI:10.1002/ps.3525

Boris Bogdanow, Henry Weisbach, Jens von Einem, Sarah Straschewski, Sebastian Voigt, Michael Winkler, Christian Hagemeier, Lüder Wiebusch: *Human cytomegalovirus tegument protein pp150 acts as a cyclin A2-CDK-dependent sensor of the host cell cycle and differentiation state*. Proceedings of the National Academy of Sciences 10/2013; 110(43)., DOI:10.1073/pnas.1312235110

9. ACKNOWLEDGMENTS

I consider myself very lucky having worked in Matthias Selbach's lab and under his supervision. I thank him very much for being always optimistic about and interested in the research I was doing. After working in his lab for all these years there was not a single event where Matthias made a decision that I was unhappy with, both personally and professionally.

I thank my dissertation committee members at the Technical University Berlin, Prof. Kurreck and Prof. Neubauer for reading this thesis and for taking part in my defense.

I am grateful to Thorsten Wolff for giving me the opportunity to work in his lab with his very friendly and welcoming members. I thank Anne, Jessi, Christin, Daniel, Marlena, Gudrun, Sandra and Matthias for taking on the challenge of maneuvering my access to the RKI and for sharing all the little details about practical work with IAV. A big thanks goes to Gudrun, who helped me a lot with the practical work and accepted with calm when I time after time misplaced something in the lab.

I thank Irmtraud Meyer. We always had very interesting, enjoyable and unique scientific conversations.

A special thanks goes to Lüder Wiebusch. Often, when I did not know how to go on with something, I called him or just went by his lab. Lüder was always a supportive friend to me. I am greatly indebted to him because of all the advice, his practical and theoretical support (and coffee) that he provided. His support greatly contributed to the outcome of the thesis. A thanks goes also out to the lab members of LPMB, Barbara, Iris and Quangh.

It is a great pleasure to thank the people who were and are working in the Selbach lab. I am grateful to all the people who have given me so much practical insight, especially Erik, Florian, Jimmy, Kamila, Djordje, Marie, Koshi-san and Katrina. Also a big thanks to Tommaso, Trendelina, Joao, Carlos, Teresa, Christopher, Amir, Damir, Benjamin and Matthias. I enjoyed working and spending quality time with all of them!

A special thanks goes to Henrik from the Selbach lab, who undertook a one year excursion into peptide modifications with me, which was a lot of fun. I am also grateful to Martha and Christian who helped me with everything (not an overstatement) from mass-spec to cloning to western blots to PCR, etc. A sad thank you goes out to Katrin Eichelbaum, who started the project and deserves a lot of credit for it. Her tragic story and the strength how she dealt with it will never be forgotten. I thank Petra and Sabine, who made MDC bureaucracy unbelievably easy for me.

A very big thanks goes to Xi Wang and Jingyi Hou, who were working at Wei Chen's lab at the MDC. They are very hard working and supportive scientists. Their help was invaluable to the success of our project. Also, I thank the Sequencing core facility at the MDC for their kind and quick service.

Finally, I am thanking my family for supporting me these years, especially my wife Katrin, who managed to support me during sad and happy times in the PhD.

10. REFERENCES

Adler, B., Sattler, C., and Adler, H. (2017). Herpesviruses and Their Host Cells: A Successful Liaison. Trends Microbiol *25*, 229-241.

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198-207.

Akarsu, H., Burmeister, W.P., Petosa, C., Petit, I., Muller, C.W., Ruigrok, R.W., and Baudin, F. (2003). Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). EMBO J *22*, 4646-4655.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H.*, et al.* (2006). The marine viromes of four oceanic regions. PLoS Biol *4*, e368.

Azab, W., Dayaram, A., Greenwood, A.D., and Osterrieder, N. (2018). How Host Specific Are Herpesviruses? Lessons from Herpesviruses Infecting Wild and Endangered Mammals. Annu Rev Virol *5*, 53-68.

Backstrom Winquist, E., Abdurahman, S., Tranell, A., Lindstrom, S., Tingsborg, S., and Schwartz, S. (2012). Inefficient splicing of segment 7 and 8 mRNAs is an inherent property of influenza virus A/Brevig Mission/1918/1 (H1N1) that causes elevated expression of NS1 protein. Virology *422*, 46-58.

Bagert, J.D., Xie, Y.J., Sweredoski, M.J., Qi, Y., Hess, S., Schuman, E.M., and Tirrell, D.A. (2014). Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture. Mol Cell Proteomics *13*, 1352-1358.

Baker, S.F., Ledwith, M.P., and Mehle, A. (2018). Differential Splicing of ANP32A in Birds Alters Its Ability to Stimulate RNA Synthesis by Restricted Influenza Polymerase. Cell Rep *24*, 2581-2588 e2584.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R.*, et al.* (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587-1593.

Batra, J., Hultquist, J.F., Liu, D., Shtanko, O., Von Dollen, J., Satkamp, L., Jang, G.M., Luthra, P., Schwarz, T.M., Small, G.I.*, et al.* (2018). Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. Cell *175*, 1917-1930 e1913.

Bercovich-Kinori, A., Tai, J., Gelbart, I.A., Shitrit, A., Ben-Moshe, S., Drori, Y., Itzkovitz, S., Mandelboim, M., and Stern-Ginossar, N. (2016). A systematic view on influenza induced host shutoff. Elife *5*.

Blight, K.J., McKeating, J.A., and Rice, C.M. (2002). Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. J Virol *76*, 13001-13014.

Bogdanow, B., Weisbach, H., von Einem, J., Straschewski, S., Voigt, S., Winkler, M., Hagemeier, C., and Wiebusch, L. (2013). Human cytomegalovirus tegument protein pp150 acts as a cyclin A2-CDK-dependent sensor of the host cell cycle and differentiation state. Proc Natl Acad Sci U S A *110*, 17510-17515.

Bogdanow, B., Zauber, H., and Selbach, M. (2016). Systematic Errors in Peptide and Protein Identification and Quantification by Modified Peptides. Mol Cell Proteomics *15*, 2791-2801.

Brauburger, K., Hume, A.J., Muhlberger, E., and Olejnik, J. (2012). Forty-five years of Marburg virus research. Viruses *4*, 1878-1927.

Bui, M., Wills, E.G., Helenius, A., and Whittaker, G.R. (2000). Role of the influenza virus M1 protein in nuclear export of viral ribonucleoproteins. J Virol *74*, 1781-1786.

Bukrinskaya, A., Starov, A., and Issayeva, C. (1981). Abortive infection of Ehrlich tumor cells by influenza virus: virus-specific RNA, protein synthesis, and transport. Intervirology *16*, 43-48.

Bullough, P.A., Hughson, F.M., Skehel, J.J., and Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. Nature *371*, 37-43.

Butt, K.M., Smith, G.J., Chen, H., Zhang, L.J., Leung, Y.H., Xu, K.M., Lim, W., Webster, R.G., Yuen, K.Y., Peiris, J.S.*, et al.* (2005). Human infection with an avian H9N2 influenza A virus in Hong Kong in 2003. J Clin Microbiol *43*, 5760-5767.

Calve, S., Witten, A.J., Ocken, A.R., and Kinzer-Ursem, T.L. (2016). Incorporation of non-canonical amino acids into the developing murine proteome. Sci Rep-Uk *6*.

Carroll, D., Daszak, P., Wolfe, N.D., Gao, G.F., Morel, C.M., Morzaria, S., Pablos-Mendez, A., Tomori, O., and Mazet, J.A.K. (2018). The Global Virome Project. Science *359*, 872-874.

Chen, J., Lee, K.H., Steinhauer, D.A., Stevens, D.J., Skehel, J.J., and Wiley, D.C. (1998). Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. Cell *95*, 409-417.

Chua, M.A., Schmid, S., Perez, J.T., Langlois, R.A., and Tenoever, B.R. (2013). Influenza A virus utilizes suboptimal splicing to coordinate the timing of infection. Cell Rep *3*, 23-29.

Cloutier, P., Poitras, C., Durand, M., Hekmat, O., Fiola-Masson, E., Bouchard, A., Faubert, D., Chabot, B., and Coulombe, B. (2017). R2TP/Prefoldin-like component RUVBL1/RUVBL2 directly interacts with ZNHIT2 to regulate assembly of U5 small nuclear ribonucleoprotein. Nat Commun *8*, 15615.

Coleman, T.P., and Roesser, J.R. (1998). RNA secondary structure: an important cis-element in rat calcitonin/CGRP pre-messenger RNA splicing. Biochemistry *37*, 15941-15950.

Cowling, B., Jin, L., Lau, E., Liao, Q., Wu, P., Jiang, H., Tsang, T., Zheng, J., and Fang, V. (2013). Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: a population-based study of laboratory-confirmed cases. The LAncet *382*, 129-137.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res *10*, 1794-1805.

Cros, J.F., Garcia-Sastre, A., and Palese, P. (2005). An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. Traffic *6*, 205-213.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics *25*, 1974-1975.

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

Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A., and Schuman, E.M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc Natl Acad Sci U S A *103*, 9482-9487.

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

Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel) *1*, 895-905.

Dubois, J., Terrier, O., and Rosa-Calatrava, M. (2014). Influenza viruses and mRNA splicing: doing more with less. MBio *5*, e00070-00014.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Eichelbaum, K., and Krijgsveld, J. (2014a). Combining pulsed SILAC labeling and click-chemistry for quantitative secretome analysis. Methods Mol Biol *1174*, 101-114.

Eichelbaum, K., and Krijgsveld, J. (2014b). Rapid temporal dynamics of transcription, protein synthesis, and secretion during macrophage activation. Mol Cell Proteomics *13*, 792-810.

Eisfeld, A.J., Neumann, G., and Kawaoka, Y. (2015). At the centre: influenza A virus ribonucleoproteins. Nat Rev Microbiol *13*, 28-41.

Falcon, A.M., Fortes, P., Marion, R.M., Beloso, A., and Ortin, J. (1999). Interaction of influenza virus NS1 protein and the human homologue of Staufen in vivo and in vitro. Nucleic Acids Res *27*, 2241-2247.

Ferhadian, D., Contrant, M., Printz-Schweigert, A., Smyth, R.P., Paillart, J.C., and Marquet, R. (2018). Structural and Functional Motifs in Influenza Virus RNAs. Front Microbiol *9*, 559.

Francis, T., and Moore, A.E. (1940). A Study of the Neurotropic Tendency in Strains of the Virus of Epidemic Influenza. J Exp Med *72*, 717-728.

Furuse, Y., Suzuki, A., Kamigaki, T., and Oshitani, H. (2009). Evolution of the M gene of the influenza A virus in different host species: large-scale sequence analysis. Virol J *6*, 67.

Gabriel, G., Dauber, B., Wolff, T., Planz, O., Klenk, H.D., and Stech, J. (2005). The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. Proc Natl Acad Sci U S A *102*, 18590-18595.

Gabriel, G., Klingel, K., Otte, A., Thiele, S., Hudjetz, B., Arman-Kalcek, G., Sauter, M., Shmidt, T., Rother, F., Baumgarte, S.*, et al.* (2011). Differential use of importin-alpha isoforms governs cell tropism and host adaptation of influenza virus. Nat Commun *2*, 156.

Galloway, S.E., Reed, M.L., Russell, C.J., and Steinhauer, D.A. (2013). Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. PLoS Pathog *9*, e1003151.

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

Garfinkel, M.S., and Katze, M.G. (1992). Translational control by influenza virus. Selective and cap-dependent translation of viral mRNAs in infected cells. J Biol Chem *267*, 9383-9390.

Garfinkel, M.S., and Katze, M.G. (1993). Translational control by influenza virus. Selective translation is mediated by sequences within the viral mRNA 5'-untranslated region. J Biol Chem *268*, 22223-22226.

Geoghegan, J.L., Duchene, S., and Holmes, E.C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog *13*, e1006215.

Geoghegan, J.L., and Holmes, E.C. (2017). Predicting virus emergence amid evolutionary noise. Open Biol *7*.

Greco, T.M., Diner, B.A., and Cristea, I.M. (2014). The Impact of Mass Spectrometry-Based Proteomics on Fundamental Discoveries in Virology. Annu Rev Virol *1*, 581-604.

Guindon, S., Delsuc, F., Dufayard, J.F., and Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol *537*, 113-137.

Halstead, S.B. (2008). Dengue virus-mosquito interactions. Annu Rev Entomol *53*, 273-291.

Harris, A., Cardone, G., Winkler, D.C., Heymann, J.B., Brecher, M., White, J.M., and Steven, A.C. (2006). Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci U S A *103*, 19123-19127.

Heikkinen, L.S., Kazlauskas, A., Melen, K., Wagner, R., Ziegler, T., Julkunen, I., and Saksela, K. (2008). Avian and 1918 Spanish influenza a virus NS1 proteins bind to Crk/CrkL Src homology 3 domains to activate host cell signaling. J Biol Chem *283*, 5719-5727.

Helenius, A. (1992). Unpacking the incoming influenza virus. Cell *69*, 577-578.

Hicks, M.J., Yang, C.R., Kotlajich, M.V., and Hertel, K.J. (2006). Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. PLoS Biol *4*, e147.

Huang, X., Zheng, M., Wang, P., Mok, B.W., Liu, S., Lau, S.Y., Chen, P., Liu, Y.C., Liu, H., Chen, Y.*, et al.* (2017). An NS-segment exonic splicing enhancer regulates influenza A virus replication in mammalian cells. Nat Commun *8*, 14751.

Hutchinson, E.C., Charles, P.D., Hester, S.S., Thomas, B., Trudgian, D., Martinez-Alonso, M., and Fodor, E. (2014). Conserved and host-specific features of influenza virion architecture. Nat Commun *5*, 4816.

Inglis, S.C. (1982). Inhibition of host protein synthesis and degradation of cellular mRNAs during infection by influenza and herpes simplex virus. Mol Cell Biol *2*, 1644-1648.

Inglis, S.C., and Brown, C.M. (1984). Differences in the control of virus mRNA splicing during permissive or abortive infection with influenza A (fowl plague) virus. J Gen Virol *65 (Pt 1)*, 153-164.

International Committee on Taxonomy of Viruses, (2019). ICTV Taxonomy Releases. https://talk.ictvonline.org/taxonomy/p/taxonomy_releases

Israel, A. (1980). Productive and abortive infection of L cells by fowl plague virus (FPV): comparison of in vivo and in vitro translation products of the virus mRNAs. J Gen Virol *47*, 473-483.

Iuliano, A.D., Roguski, K.M., Chang, H.H., Muscatello, D.J., Palekar, R., Tempia, S., Cohen, C., Gran, J.M., Schanzer, D., Cowling, B.J.*, et al.* (2018). Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet *391*, 1285-1300.

Jacobs, S.E., Lamson, D.M., St George, K., and Walsh, T.J. (2013). Human rhinoviruses. Clin Microbiol Rev *26*, 135-162.

Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K.*, et al.* (2011). Global landscape of HIV-human protein complexes. Nature *481*, 365-370.

Jagger, B.W., Wise, H.M., Kash, J.C., Walters, K.A., Wills, N.M., Xiao, Y.L., Dunfee, R.L., Schwartzman, L.M., Ozinsky, A., Bell, G.L.*, et al.* (2012). An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. Science *337*, 199-204.

Jorba, N., Coloma, R., and Ortin, J. (2009). Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. PLoS Pathog *5*, e1000462.

Killip, M.J., Fodor, E., and Randall, R.E. (2015). Influenza virus activation of the interferon system. Virus Res *209*, 11-22.

Kochs, G., Garcia-Sastre, A., and Martinez-Sobrido, L. (2007). Multiple anti-interferon actions of the influenza A virus NS1 protein. J Virol *81*, 7011-7021.

Lai, D., and Meyer, I.M. (2014). e-RNA: a collection of web servers for comparative RNA structure prediction and visualisation. Nucleic Acids Res *42*, W373-376.

Lamb, R.A., and Choppin, P.W. (1979). Segment 8 of the influenza virus genome is unique in coding for two polypeptides. Proc Natl Acad Sci U S A *76*, 4908-4912.

Lamb, R.A., Lai, C.J., and Choppin, P.W. (1981). Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. Proc Natl Acad Sci U S A *78*, 4170-4174.

Lau, S.C., and Scholtissek, C. (1995). Abortive infection of Vero cells by an influenza A virus (FPV). Virology *212*, 225-231.

Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J., Holmes, E.C., and Zhang, Y.Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife *4*.

Long, J.S., Giotis, E.S., Moncorge, O., Frise, R., Mistry, B., James, J., Morisson, M., Iqbal, M., Vignal, A., Skinner, M.A.*, et al.* (2016). Species difference in ANP32A underlies influenza A virus polymerase host restriction. Nature *529*, 101-104.

Long, J.S., Mistry, B., Haslam, S.M., and Barclay, W.S. (2019). Host and viral determinants of influenza A virus species specificity. Nat Rev Microbiol *17*, 67-81.

Lowe, A.B. (2014). Thiol-yne 'click'/coupling chemistry and recent applications in polymer and materials synthesis and modification. Polymer *55*, 5517-5549.

Luo, G.X., Luytjes, W., Enami, M., and Palese, P. (1991). The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure. J Virol *65*, 2861-2867.

Lustig, A., and Levine, A.J. (1992). One hundred years of virology. J Virol *66*, 4629-4631.

Mair, C.M., Ludwig, K., Herrmann, A., and Sieben, C. (2014). Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection. Biochim Biophys Acta *1838*, 1153-1168.

Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006). Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. J Am Soc Mass Spectrom *17*, 977-982.

Malvy, D., McElroy, A.K., de Clerck, H., Gunther, S., and van Griensven, J. (2019). Ebola virus disease. Lancet *393*, 936-948.

Man, O., and Pilpel, Y. (2007). Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species (vol 39, pg 415, 2007). Nat Genet *39*, 688-688.

Manz, B., Dornfeld, D., Gotz, V., Zell, R., Zimmermann, P., Haller, O., Kochs, G., and Schwemmle, M. (2013). Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. PLoS Pathog *9*, e1003279.

Marjuki, H., Alam, M.I., Ehrhardt, C., Wagner, R., Planz, O., Klenk, H.D., Ludwig, S., and Pleschka, S. (2006). Membrane accumulation of influenza A virus hemagglutinin triggers nuclear export of the viral genome via protein kinase Calpha-mediated activation of ERK signaling. J Biol Chem *281*, 16707-16715.

Martin, K., and Helenius, A. (1991a). Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. Cell *67*, 117-130.

Martin, K., and Helenius, A. (1991b). Transport of incoming influenza virus nucleocapsids into the nucleus. J Virol *65*, 232-244.

Matrosovich, M.N., Gambaryan, A.S., Teneberg, S., Piskarev, V.E., Yamnikova, S.S., Lvov, D.K., Robertson, J.S., and Karlsson, K.A. (1997). Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. Virology *233*, 224-234.

Matthaei, M., Budt, M., and Wolff, T. (2013). Highly pathogenic H5N1 influenza A virus strains provoke heterogeneous IFN-alpha/beta responses that distinctively affect viral propagation in human cells. PLoS One *8*, e56659.

McShane, E., Sin, C., Zauber, H., Wells, J.N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J.A.*, et al.* (2016). Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. Cell *167*, 803-815 e821.

Mehle, A., and Doudna, J.A. (2008). An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. Cell Host Microbe *4*, 111-122.

Meyer, K., and Selbach, M. (2015). Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. Front Genet *6*, 237.

Mor, A., White, A., Zhang, K., Thompson, M., Esparza, M., Munoz-Moreno, R., Koide, K., Lynch, K.W., Garcia-Sastre, A., and Fontoura, B.M. (2016). Influenza Virus mRNA Trafficking Through Host Nuclear Speckles. Nat Microbiol *2016*.

Morens, D.M., Taubenberger, J.K., and Fauci, A.S. (2009). The persistent legacy of the 1918 influenza virus. N Engl J Med *361*, 225-229.

Morwitzer, M.J., Tritsch, S.R., Cazares, L.H., Ward, M.D., Nuss, J.E., Bavari, S., and Reid, S.P. (2019). Identification of RUVBL1 and RUVBL2 as Novel Cellular Interactors of the Ebola Virus Nucleoprotein. Viruses *11*.

Moss, W.N., Dela-Moss, L.I., Kierzek, E., Kierzek, R., Priore, S.F., and Turner, D.H. (2012). The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. PLoS One *7*, e38323.

Nayak, D.P., Hui, E.K., and Barman, S. (2004). Assembly and budding of influenza virus. Virus Res *106*, 147-165.

Nemeroff, M.E., Barabino, S.M., Li, Y., Keller, W., and Krug, R.M. (1998). Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3'end formation of cellular pre-mRNAs. Mol Cell *1*, 991-1000.

Neumann, G., and Kawaoka, Y. (2015). Transmission of influenza A viruses. Virology *479-480*, 234-246.

Neumann, G., Noda, T., and Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature *459*, 931-939.

Nielsen, M.L., Savitski, M.M., and Zubarev, R.A. (2006). Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. Mol Cell Proteomics *5*, 2384-2391.

Nilsson, T., Mann, M., Aebersold, R., Yates, J.R., 3rd, Bairoch, A., and Bergeron, J.J. (2010). Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Methods *7*, 681-685.

O'Neill, R.E., Talon, J., and Palese, P. (1998). The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. EMBO J *17*, 288-296.

Ogoina, D., and Onyemelukwe, G.C. (2009). The role of infections in the emergence of non-communicable diseases (NCDs): Compelling needs for novel strategies in the developing world. J Infect Public Health *2*, 14-29.

Olsen, B., Munster, V.J., Wallensten, A., Waldenstrom, J., Osterhaus, A.D.M.E., and Fouchier, R.A.M. (2006). Global patterns of influenza A virus in wild birds. Science *312*, 384-388.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics *1*, 376-386.

Oxford Dictionaries, (2019). Defintion of virus in English. *2019*, https://en.oxforddictionaries.com/definition/virus

Palese, P., Tobita, K., Ueda, M., and Compans, R.W. (1974). Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. Virology *61*, 397-410.

Pedersen, J.S., Meyer, I.M., Forsberg, R., Simmonds, P., and Hein, J. (2004). A comparative method for finding and folding RNA secondary structures within protein-coding regions. Nucleic Acids Res *32*, 4925-4936.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M.*, et al.* (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res *47*, D442-D450.

Pfeiffer, R. (1892). I.-Preliminary Communication on the Exciting causes of Influenza. Br Med J *1*, 128.

Pichlmair, A., Kandasamy, K., Alvisi, G., Mulhern, O., Sacco, R., Habjan, M., Binder, M., Stefanovic, A., Eberle, C.A., Goncalves, A.*, et al.* (2012). Viral immune

modulators perturb the human molecular network by common and unique strategies. Nature *487*, 486-490.

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

Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L., and Lloyd-Smith, J.O. (2017). Pathways to zoonotic spillover. Nat Rev Microbiol *15*, 502-510.

Ramage, H.R., Kumar, G.R., Verschueren, E., Johnson, J.R., Von Dollen, J., Johnson, T., Newton, B., Shah, P., Horner, J., Krogan, N.J.*, et al.* (2015). A combined proteomics/genomics approach links hepatitis C virus infection with nonsense-mediated mRNA decay. Mol Cell *57*, 329-340.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem *75*, 663-670.

Rehwinkel, J., Tan, C.P., Goubau, D., Schulz, O., Pichlmair, A., Bier, K., Robb, N., Vreede, F., Barclay, W., Fodor, E.*, et al.* (2010). RIG-I detects viral genomic RNA during negative-strand RNA virus infection. Cell *140*, 397-408.

Reid, A.H., Fanning, T.G., Janczewski, T.A., McCall, S., and Taubenberger, J.K. (2002). Characterization of the 1918 "Spanish" influenza virus matrix gene segment. J Virol *76*, 10717-10723.

Resa-Infante, P., Jorba, N., Zamarreno, N., Fernandez, Y., Juarez, S., and Ortin, J. (2008). The host-dependent interaction of alpha-importins with influenza PB2 polymerase subunit is required for virus RNA replication. PLoS One *3*, e3904.

Rivas, H.G., Schmaling, S.K., and Gaglia, M.M. (2016). Shutoff of Host Gene Expression in Influenza A Virus and Herpesviruses: Similar Mechanisms and Common Themes. Viruses *8*, 102.

Robb, N.C., and Fodor, E. (2012). The accumulation of influenza A virus segment 7 spliced mRNAs is regulated by the NS1 protein. J Gen Virol *93*, 113-118.

Rodriguez, A., Perez-Gonzalez, A., and Nieto, A. (2007). Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. J Virol *81*, 5315-5324.

Roizmann, B., Desrosiers, R.C., Fleckenstein, B., Lopez, C., Minson, A.C., and Studdert, M.J. (1992). The family Herpesviridae: an update. The Herpesvirus Study Group of the International Committee on Taxonomy of Viruses. Arch Virol *123*, 425-449.

Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelmant, G., Calderwood, M.A., Rolland, T., Grace, M., Dricot, A., Askenazi, M., Tavares, M.*, et al.* (2012). Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature *487*, 491-495.

Sadewasser, A., Paki, K., Eichelbaum, K., Bogdanow, B., Saenger, S., Budt, M., Lesch, M., Hinz, K.P., Herrmann, A., Meyer, T.F.*, et al.* (2017). Quantitative Proteomic Approach Identifies Vpr Binding Protein as Novel Host Factor Supporting Influenza A Virus Infections in Human Cells. Mol Cell Proteomics *16*, 728-742.

Saunders-Hastings, P.R., and Krewski, D. (2016). Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission. Pathogens *5*.

Scaturro, P., Stukalov, A., Haas, D.A., Cortese, M., Draganova, K., Plaszczyca, A., Bartenschlager, R., Gotz, M., and Pichlmair, A. (2018). An orthogonal proteomic survey uncovers novel Zika virus host factors. Nature *561*, 253-257.

Schlesinger, R.W., Bradshaw, G.L., Barbone, F., Reinacher, M., Rott, R., and Husak, P. (1989). Role of hemagglutinin cleavage and expression of M1 protein in replication of A/WS/33, A/PR/8/34, and WSN influenza viruses in mouse brain. J Virol *63*, 1695-1703.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671-675.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Schwanhausser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009). Global analysis of cellular protein translation by pulsed SILAC. Proteomics *9*, 205-209.

Selman, M., Dankar, S.K., Forbes, N.E., Jia, J.J., and Brown, E.G. (2012). Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. Emerg Microbes Infect *1*, e42.

Shah, P.S., Link, N., Jang, G.M., Sharp, P.P., Zhu, T., Swaney, D.L., Johnson, J.R., Von Dollen, J., Ramage, H.R., Satkamp, L.*, et al.* (2018). Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. Cell *175*, 1931-1945 e1918.

Shapiro, G.I., Gurney, T., Jr., and Krug, R.M. (1987). Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. J Virol *61*, 764-773.

Shi, M., Lin, X.D., Chen, X., Tian, J.H., Chen, L.J., Li, K., Wang, W., Eden, J.S., Shen, J.J., Liu, L.*, et al.* (2018). The evolutionary history of vertebrate RNA viruses. Nature *556*, 197-202.

Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S.*, et al.* (2016). Redefining the invertebrate RNA virosphere. Nature *540*, 539-543.

Shih, S.R., and Krug, R.M. (1996). Novel exploitation of a nuclear function by influenza virus: the cellular SF2/ASF splicing factor controls the amount of the essential viral M2 ion channel protein in infected cells. EMBO J *15*, 5415-5427.

Shih, S.R., Nemeroff, M.E., and Krug, R.M. (1995). The choice of alternative 5' splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex. Proc Natl Acad Sci U S A *92*, 6324-6328.

Shih, S.R., Suen, P.C., Chen, Y.S., and Chang, S.C. (1998). A novel spliced transcript of influenza A/WSN/33 virus. Virus Genes *17*, 179-183.

Shope, R.E. (1931). The Etiology of Swine Influenza. Science *73*, 214-215.

Smith, G.J., Vijaykrishna, D., Bahl, J., Lycett, S.J., Worobey, M., Pybus, O.G., Ma, S.K., Cheung, C.L., Raghwani, J., Bhatt, S.*, et al.* (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature *459*, 1122-1125.

Smith, G.L., Levin, J.Z., Palese, P., and Moss, B. (1987). Synthesis and cellular location of the ten influenza polypeptides individually expressed by recombinant vaccinia viruses. Virology *160*, 336-345.

Smith, W., Andrewes, C.H., and Laidlaw, P.P. (1995). A virus obtained from influenza patients (Reprinted from Lancet, July 8, pg 66-68, 1933). Rev Med Virol *5*, 187-191.

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

Stern-Ginossar, N., Thompson, S.R., Mathews, M.B., and Mohr, I. (2019). Translational Control in Virus-Infected Cells. Cold Spring Harb Perspect Biol *11*.

Tan, S.L., and Katze, M.G. (1998). Biochemical and genetic evidence for complex formation between the influenza A virus NS1 protein and the interferon-induced PKR protein kinase. J Interferon Cytokine Res *18*, 757-766.

Taubenberger, J.K., and Kash, J.C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe *7*, 440-451.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem *75*, 1895-1904.

Thompson, M.G., Munoz-Moreno, R., Bhat, P., Roytenberg, R., Lindberg, J., Gazzara, M.R., Mallory, M.J., Zhang, K., Garcia-Sastre, A., Fontoura, B.M.A.*, et al.* (2018). Co-regulatory activity of hnRNP K and NS1-BP in influenza and human mRNA splicing. Nat Commun *9*, 2407.

Tran, T.H., Nguyen, T.L., Nguyen, T.D., Luong, T.S., Pham, P.M., Nguyen v, V., Pham, T.S., Vo, C.D., Le, T.Q., Ngo, T.T.*, et al.* (2004). Avian influenza A (H5N1) in 10 patients in Vietnam. N Engl J Med *350*, 1179-1188.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and

quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Tripathi, S., Pohl, M.O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D.A., Moulton, H.M., DeJesus, P., Che, J., Mulder, L.C.*, et al.* (2015). Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. Cell Host Microbe *18*, 723-735.

Tsai, P.L., Chiou, N.T., Kuss, S., Garcia-Sastre, A., Lynch, K.W., and Fontoura, B.M. (2013). Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. PLoS Pathog *9*, e1003460.

Valcarcel, J., Portela, A., and Ortin, J. (1991). Regulated M1 mRNA splicing in influenza virus-infected cells. J Gen Virol *72 (Pt 6)*, 1301-1308.

Wagner, R., Matrosovich, M., and Klenk, H.D. (2002). Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Rev Med Virol *12*, 159-166.

Watakabe, A., Inoue, K., Sakamoto, H., and Shimura, Y. (1989). A secondary structure at the 3' splice site affects the in vitro splicing reaction of mouse immunoglobulin mu chain pre-mRNAs. Nucleic Acids Res *17*, 8159-8169.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature *460*, 711-716.

Weber, M., Sediri, H., Felgenhauer, U., Binzen, I., Banfer, S., Jacob, R., Brunotte, L., Garcia-Sastre, A., Schmid-Burgk, J.L., Schmidt, T.*, et al.* (2015). Influenza virus adaptation PB2-627K modulates nucleocapsid inhibition by the pathogen sensor RIG-I. Cell Host Microbe *17*, 309-319.

Wessel, D., and Flugge, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem *138*, 141-143.

Wigington, C.H., Sonderegger, D., Brussaard, C.P., Buchan, A., Finke, J.F., Fuhrman, J.A., Lennon, J.T., Middelboe, M., Suttle, C.A., Stock, C.*, et al.* (2016). Re-examination of the relationship between marine virus and microbial cell abundances. Nat Microbiol *1*, 15024.

Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., Kang, Z.H., Robb, N., Schwartzman, L.M., Kash, J.C., Fodor, E., Firth, A.E.*, et al.* (2012). Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. PLoS Pathog *8*, e1002998.

Woolhouse, M., Scott, F., Hudson, Z., Howey, R., and Chase-Topping, M. (2012). Human viruses: discovery and emergence. Philos Trans R Soc Lond B Biol Sci *367*, 2864-2871.

World Health Organization, (2019a). Fact sheet seasonal Influenza. https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal

World Health Organization, (2019b). Monthly Risk Assessment. *2019*, https://www.who.int/influenza/human_animal_interface/HAI_Risk_Assessment/en/

Worobey, M., Han, G.Z., and Rambaut, A. (2014a). Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. Proc Natl Acad Sci U S A *111*, 8107-8112.

Worobey, M., Han, G.Z., and Rambaut, A. (2014b). A synchronized global sweep of the internal genes of modern avian influenza virus. Nature *508*, 254-257.

Yasunaga, A., Hanna, S.L., Li, J., Cho, H., Rose, P.P., Spiridigliozzi, A., Gold, B., Diamond, M.S., and Cherry, S. (2014). Genome-wide RNAi screen identifies broadly-acting host factors that inhibit arbovirus infection. PLoS Pathog *10*, e1003914.

Yue, J., Lai, F., Beckedorff, F., Zhang, A., Pastori, C., and Shiekhattar, R. (2017). Integrator orchestrates RAS/ERK1/2 signaling transcriptional programs. Genes Dev *31*, 1809-1820.

Zhang, Y., Aevermann, B.D., Anderson, T.K., Burke, D.F., Dauphin, G., Gu, Z., He, S., Kumar, S., Larsen, C.N., Lee, A.J.*, et al.* (2017). Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res *45*, D466-D474.

Zhou, T., Gu, W., Ma, J., Sun, X., and Lu, Z. (2005). Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. Biosystems *81*, 77-86.