Identification of *cis*-Acting Elements that Regulate Influenza A Virus Segment 7 Differential mRNA Splicing

Shu Zhou, Rute Maria Pinto, Hou Wei Chook, Hui Min Lee, Eleanor R Gaunt, Paul Digard

The Roslin Institute, University of Edinburgh, EH25 9RG, Edinburgh, United Kingdom

Short title: Regulation of influenza A virus segment 7 splicing

Key words: influenza A virus, matrix, M1, M2, ion channel, splicing

Corresponding author: Paul Digard

Email: Paul.digard@roslin.ed.ac.uk

Abstract

Influenza A virus segment 7 encodes the M1 matrix protein and M2 ion channel from an unspliced primary transcript mRNA1 and spliced mRNA2, respectively. Both proteins are essential for efficient virus replication, but the intrinsic sequences that determine segment 7 splicing balance are not fully defined. Here, we systematically investigated how conserved *cis*-acting splicing elements regulate 7 splicing. We found that while the splice acceptor sequence is optimal, both potential upstream polypyrimidine tracts regulate splicing, the 5' region acting positively and the 3' region negatively to maintain M1 expression. We could not unambiguously identify the mRNA2 branch point, but two closely spaced motifs can act redundantly. Saturation mutagenesis of the mRNA2 splice donor sequence showed that it is nearly optimal but that nucleotides 51 and 55 modestly downregulate splicing efficiency. Most introduced mutations that change splicing balance are nonsynonymous in M1. Nevertheless, mutations in the polypyrimidine tract and branch points that significantly perturbed M1:M2 balance resulted in viable viruses. In contrast, the splice donor sequence was intolerant of mutation. We conclude that M1:M2 balance is maintained via suboptimal splice donor and polypyrimidine tract splicing signals, with the latter allowing more evolutionary flexibility to vary the ratio and/or M1 sequence.

Introduction

Influenza A virus (IAV) is an enveloped negative-sense single-stranded RNA virus with eight genome segments. Ten proteins encoded by these segments are essential for virus fitness outside of the laboratory: segments 1-6 encode polymerase basic subunit 1 (PB1), polymerase basic subunit 2 (PB2), polymerase acidic subunit (PA), haemagglutinin (HA), nucleoprotein (NP), and neuraminidase (NA). Segments 7 and 8 encode more than one protein by alternative splicing: matrix protein (M1), matrix 2 ion channel protein (M2) from segment 7, nonstructural protein (NS1) and nuclear export protein (NEP) from segment 8 (1). The viral proteome is further elaborated through a variety of transcriptional and translational mechanisms to produce a virus strain-dependent suite of accessory protein species (2).

The M1 matrix protein of the virus has multiple functions. As well as forming a helical filament under the virus envelope, it recruits several other viral components to assemble at the site of budding (3, 4) and determines viral particle morphology (5, 6). M1 also binds to viral ribonucleoproteins (RNPs) in a regulated manner to control their nuclear import and export (7). The M2 protein is a 97 aa proton channel that is translated from the main spliced transcript of segment 7 mRNA and that mediates acidification of the interior of the virus particle in endosomes when the virus enters a cell (8, 9). It has also been proposed to mediate membrane scission of budding virions (10) as well as to block autophagosome fusion with lysosomes (11). Although both M1 and M2 are found in the virion, the copy number of M2 is two orders of magnitude lower than of M1 (12). Despite the markedly different demands of M1 and M2 for the virion particle, expression levels are more equal in cells, with M2 comprising between 10% to 60% of segment 7 protein production, depending on the cell and virus strain (13), implying M1 and M2 expression is regulated. Primarily, this is thought to occur at the level of mRNA splicing. Newly transcribed segment 7 mRNA

molecules accumulate in nuclear speckles for splicing at the earlier stages of infection and although measurements of the ratio of mRNA2:mRNA1 varies according to time-point, virus strain, host cell and assay method, the unspliced transcript tends to predominate (13-16).

Segment 7 mRNA splicing is more complex than simply producing mRNA2. At least four spliced transcripts have been identified (17-20) in addition to unspliced mRNA1. The spliced mRNAs use different splice donors (SD) but share the same splice acceptor site (SA). Spliced mRNA2 encodes the M2 ion channel protein (8). Spliced mRNA3 is not required for virus replication in tissue culture and no encoded protein has yet been identified (21). mRNA4 encodes an M2 variant ("M42") with a distinct ectodomain but identical C-terminal domain (22). All IAV strains produce unspliced mRNA1 and mRNA2, but the abundance of mRNA3 is variable among strains and only a small minority of strains produce mRNA4 (19, 22).

Four *cis*-acting elements are required for cellular pre-mRNA splicing: an SD at the 5'-end of the intron/exon boundary; and three separate but usually juxtaposed elements: a branch point (BP) and polypyrimidine tract (PPT) at the 3’end of the intron and a 3’-SA sequence overlapping the intron/exon boundary (Figure 1A). Most introns have highly conserved ends: a GU dinucleotide at the 5’end and an AG dinucleotide at the 3’ end (23-25). These lie within broader consensus sequences for SD and SA sites, which are MAG|GURAGU and YAG| respectively (26-28). The BP consensus (YUN<u>A</u>Y, with the underlined A being the actual branch point) is more loosely conserved, while the PPT is usually a stretch that contains around 10 pyrimidines located 5-40 nucleotides upstream of the SA (29). The SD sequence is recognised by U1 snRNP through RNA-RNA base pairing to initiate splicing (26, 30). After cleaving at the 5’ end of the intron, a guanine at the 5'-end conjugates to the BP adenine to form a lariat structure (29). The two U2 auxiliary factor (U2AF) subunits recognise the PPT and SA, respectively, recruiting U2 snRNP and

promoting spliceosome assembly (31, 32). After that, the two exons ligate and the intron in lariat configuration is excised and degraded (29).

The cellular splicing machinery recognises the *cis*-acting splicing elements based on their match to the conserved consensus sequences. Since the cellular spliceosome mediates IAV segment 7 mRNA splicing (33-35), the splicing signals in IAV mRNA generally, though imperfectly, resemble those of the host cells. However, in contrast to cellular mRNA, both unspliced and spliced transcripts of segment 7 are translated to essential viral proteins. Therefore, IAV must incompletely splice segment 7 to ensure correctly balanced M1 and M2 production. One possibility is that viral factors ensure incomplete splicing, and there is evidence to support regulation by NS1 and/or the viral polymerase complex. The cellular proteins NS1-BP, hnRNP and ASF/SF1 have been confirmed to regulate segment 7 mRNA splicing (33, 34, 36). Also, the viral polymerase and NS1 may also be involved in segment 7 mRNA splicing (33, 34, 37, 38). However, segment 7 mRNA is incompletely spliced even when expressed by cellular RNA polymerase II in the absence of other IAV proteins (37, 39, 40). The segment 7 mRNA2 splice sites are imperfect matches for the core cellular consensus sequences (Figure 1) and from mutational studies of cellular splicing, these mismatches are likely to impair the recognition of *trans*-acting elements (28), which may result in reduced splicing efficiency. However, splice site efficiency is difficult to predict from sequence (28). Other uncertainties are that the segment 7 BP has never been experimentally mapped, only computationally predicted, while the extent and location of the PPT also remain unclear (41).

In this study, we set out to test the hypothesis that the primary balance between M1 and M2 mRNAs is set by the conserved *cis*-acting splice site elements of IAV segment 7 splicing, by systematically characterising the strength and location of the SD, BP, PPT and SA motifs. We find that segment 7 has a strong SA and redundant BP sequences, and that the

key suboptimal splicing signals that give the desired mRNA1/M1 and mRNA2/M2 ratios are the SD and PPT signals.

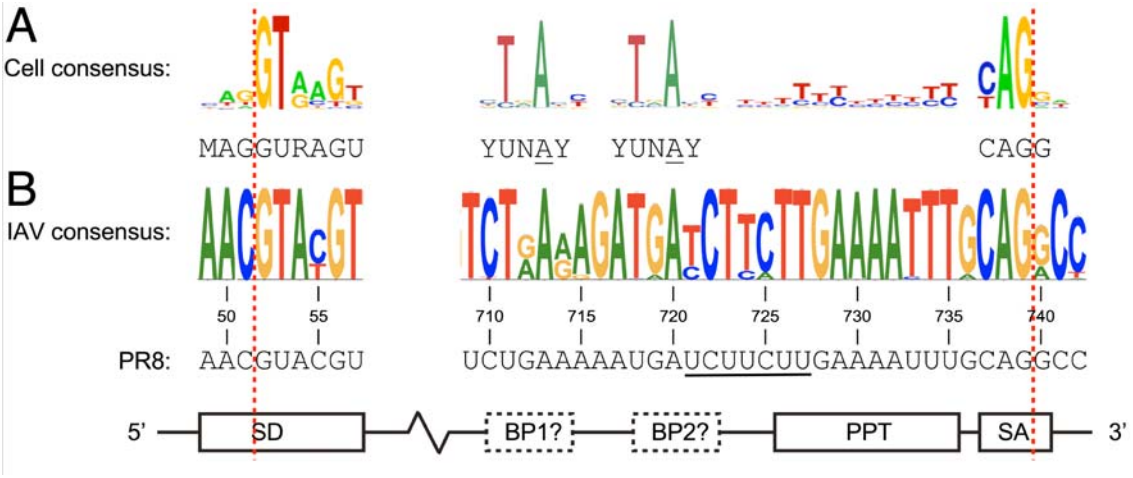

Figure 1. Consensus alignments of cellular and IAV segment 7 *cis*-acting splice elements. Sequence logos for (A) a selection of experimentally validated mammalian mRNAs (SD, PPT and SA motifs, adapted with permission from (42, 43); BP motifs adapted from (44) under CC BY 4.0 permission) and (B) IAV segment 7 sequences are shown, along with the cellular consensus (M indicates A or C and Y indicates C or U/T). The sequence of the A/Puerto Rico/8/34 (PR8) virus strain is also shown. Note that the sequence logos were generated from sequences downloaded from NCBI and thus follow the convention of T rather than U for RNA. Red dashed lines indicate splice sites (boundary between intron and exon). Boxes indicate splice site elements. SD: Splice donor; BP: Branch point; PPT: Polypyrimidine Tract (following the cellular consensus location and length - the underline region from nucleotides 721-727 indicates an alternative prediction (41)); SA: Splice Acceptor. Dashed boxes indicate that the location of IAV segment 7 BP is unclear.

Materials and Methods

Bioinformatics analyses

To generate a consensus sequence for IAV segment 7 mRNA2 splice sites, 13,280 complete IAV segment 7 sequences (all subtypes and host species) were downloaded from the National Centre for Biotechnology Information (NCBI). The viral sequence logos were generated

using WebLogo (42, 43). Counterpart sequence logos representing the consensus SD, SA and PPT sequences from experimentally verified human genes, originally abstracted from the Exon-Intron database (45), were adapted with permission from https://weblogo.berkeley.edu/examples.html. The consensus BP sequence from the bovine genome was modified with permission (https://creativecommons.org/licenses/by/4.0/) from reference (44). To predict potential splice site motifs in segment 7, the A/Puerto Rico/8/1934 (PR8) segment 7 sequence (GeneBank accession: EF467824) was truncated at the polyadenylation signal (to mimic the mRNA sequence) and used as input for various published prediction programmes: Alternative Splice Site Predictor (ASSP) (46), GENSCAN (47), Human Splicing Finder (HSF) (48), NetGene2 (49, 50), NNSPLICE (51), Spliceator (52) and SVM-BPFinder (53). Default parameters were used, other than for HSF, where HSF scoring was used, and SVM-BP, where the maxlen parameter was increased to 1000 to scan the entire segment. In addition, the output scores from SVM-BP were converted to "confidence values" (0-1) using the maximum and minimum scores for cellular genes given in (53) (approximately +5 to - 3.5 for BP sites and 0 to 31 for PPT tracts in introns of < 1000 ntds).

Cells, viruses and plasmids

Madin-Darby canine kidney (MDCK) and Human embryonic kidney cells (HEK 293T) were cultured according to standard procedures (54). Cells were routinely tested for mycoplasma contamination. Wild type (WT) and modified viruses were rescued by eight plasmid transfections (55) into HEK 293T cells using Lipofectamine 2000 following the manufacturer's instructions. Then, the virus was propagated in MDCK cells as previously described (54). Site-directed mutagenesis of segment 7 was performed using Q5 Site-Directed Mutagenesis Kits (New England Biolabs, E0554) following the manufacturer's

instructions. All constructs were confirmed by Sanger sequencing (GENEWIZ). To titre viruses, confluent monolayers of MDCK cells were infected with serially-diluted viral samples and incubated for one hour, followed by Avicel overlay and toluidine blue staining. To quantify plaque size, fixed plates were scanned in an Epson Perfection V750 Pro Scanner. at a resolution of 360 dpi. Images were converted to grey scale, contrast adjusted and the area of each plaque determined in ImageJ 2. 5- 20 plaques were selected and measured per experiment and normalised to the values from WT virus. For protein and mRNA analysis in single- cycle infection, cells were infected at an MOI of 5.

Virion RNA extraction and sequencing

Viral genomic RNA was extracted from virus stocks using a QIAamp Viral RNA Kit according to the manufacturer's instructions and DNase-treated using TURBO DNA-free kits (Thermo Fisher Scientific). Reverse transcription was performed using Superscript IV (Thermo Fisher Scientific) and Uni 12 primer (AGCAAAAGCAGG). cDNA was used as a template for PCR using PrimeSTAR GXL Premix (Takara Bio Inc). Forward and reverse primers AGCAAAAGCAGGTAGATATTG and AGTAGAAACAAGGTAGTTTTTTAC respectively were used to amplify a segment 7 amplicon. PCR conditions were 98°C for 5 min, 35 cycles of 98°C for 10s, 55°C for 15s and 68°C for 2 min. PCR products were purified by Monarch DNA Gel Extraction Kit (New England Biolabs) and analysed by Sanger sequencing (GENEWIZ).

Minireplicon assay

HEK 293T cells were seeded in 24-well plates and transfected the next day with 50 ng of the desired combinations of PR8 reverse genetics plasmids expressing combinations of PB2, PB1, PA, NP, NS and M segments using lipofectamine 2000 (Thermo Fisher Scientific) following the manufacturer's instructions. The empty vector plasmid containing bi-directional RNA PolI and PolII promoters ("pDUAL") (55) was used as a negative control by replacing the segment 7 plasmid. The medium was aspirated 48 hours after transfection, and cells were lysed using Laemmli buffer (20% glycerol, 2% Sodium Dodecyl Sulfate (SDS), 100 mM Dithiothreitol (DTT), 24 mM Tris pH 6.8, 0.016% bromophenol blue) for western blotting or lysed for total RNA extraction as described below.

RT-PCR and splicing analysis

Total RNA was extracted from virus-infected or plasmid-transfected MDCK or 293T cells using the RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. After DNA treatment using a TURBO DNA-free Kit (Thermo Fisher Scientific), cDNA was generated using Lunascript RT Master Mix Kit (New England Biolabs) and oligo dT (IDT) as a primer. The cDNA was amplified using PrimeStar Max Premix (Takara Bio Inc.). M segment PCR was performed with the following primers: forward primer: GAGTCTTCTAACCGAGGTC for constructs with SD mutations or TTCTAACCGAGGTCGAAAC for the rest of the analyses and a common reverse primer: CCACAGCACTCTGCTGTTCCT. mRNA1 and mRNA2 were distinguished based on the sizes of their PCR products: 939 or 934 bp for mRNA1 and 251 or 246 bp for mRNA2, depending on the forward primer used. NP mRNA PCR was performed with forward primer: ACGGAAAGTGGATGAGAGAACTCAT and reverse primer: CAACCGACCCTCTCAATATGAGTGC. PCR conditions were 98 °C for 5 min, 23-30

cycles of 98 °C for 10s, 55 °C for 15s and 68 °C for 10s. PCR products were separated in 2% agarose gels and imaged using an Odyssey Fc imaging system (Li-Cor Inc.) after staining with SYBR™ Safe DNA Gel Stain (Thermo Fisher).

Western blotting analysis

SDS-PAGE followed by western blotting was performed using the Trans-Blot Transfer System (BioRad) according to the manufacturer's instructions. Membranes were blocked with phosphate-buffered saline (PBS)/0.1% Tween20/5% skimmed milk at room temperature for one hour, washed with PBS/0.1% Tween20 and incubated with primary antibody diluted in blocking media. Following overnight incubation at 4 °C, membranes were washed with PBS/0.1% Tween20 and incubated with fluorophore-labelled secondary antibody diluted in blocking media for one hour. Membranes were imaged using an Odyssey Fc imaging system (Li-Cor Inc.). Image acquisition and analysis were performed using Image Studio Lite software. Purchased monoclonal antibodies were against α-tubulin (Thermo Fisher Scientific, Clone YL1/2), IAV M2 (Abcam, Clone 14C2), and IAV NP (Abcam, Clone C43). Purchased polyclonal antibodies were against IAV M1 (GeneTex, GTX125928) and M2 (Thermo Fisher Scientific, PA5-32233). In-house polyclonal rabbit anti-NP (antiserum A2915) and anti-NS1 (antiserum V29) have been previously described (56). Purchased secondary antibodies were IRDye 800CW anti-Mouse (Li-Cor Inc, 925-32212), IRDye 680RD anti-Rat (Li-Cor Inc, 926-68076) and IRDye 800CW anti-Rabbit (Li-Cor Inc, 926-32213). Densitometric analysis of western blots was done using the gel plot function in ImageJ 2 (57).

Statistical Analyses

All experiments that were statistically analysed were performed in at least biological triplicates. All graphical data were plotted and analysed using GraphPad Prism 10. Statistical tests used an ordinary One-way ANOVA test. Multiple comparisons were performed against PR8 WT using Dunnett's post-test. **** = $P<0.0001$; *** = $P<0.001$; ** = $P<0.01$; * = $P<0.05$; absence of significance symbols = $P>0.05$. Error bars indicate the standard error of the mean.

Results

Bioinformatic analysis of segment 7 splicing *cis*-acting elements

Splicing of cellular mRNAs is directed by well-defined but degenerate consensus sequences (28). Outside of the almost obligatory GU and AG nucleotides at either end of the intron, individual pre-mRNA sequences are rarely complete matches to the consensus (Figure 1A) and this can affect splice site selection (58). The SD sequence of IAV segment 7 mRNA2 is a reasonable match to the cellular consensus, possessing the obligatory intron boundary dinucleotide as well as matching at -3, +3 and +4, +5 positions (Figure 1B). The SA site is a better match to consensus; perfect in many strains and varying only at the +1 position in a minority of isolates. The viral PPT has been suggested to be composed of a stretch of 7 pyrimidines running from position 721 - 727 (41), but this would be both shorter and spaced further away from the SA than the norm (59). The region immediately 5' to the SA site where cellular PPTs are usually found is not especially pyrimidine-rich, having a run of 5 purines upstream of three uracils (Figure 1B). A conserved adenosine at position 720 has been predicted to be the segment 7 BP (41), based on its contextual match to the cellular consensus. However, A713 is also in a similar context, only deviating from consensus at the 3'-end of the motif. Thus overall, the segment 7 splice site motifs match the cellular consensus to varying degrees.

Tool	SD	BP1 (A713)	BP2 (A720)	PPT	SA
ASSP	0.94				0.46
Genscan	X				X
HSF	0.81	0.83	X		0.79
NetGene2	0.0				X
NNSPLICE	0.98				0.24
Spliceator	0.94				0.92
SVM-BPFinder		0.42	0.39	0.61	

Table 1. Splice site prediction scores for PR8 segment 7. Scores are confidence values (0 to 1). Blanks indicate that the tool does not predict a particular splice signal, crosses (X) indicate that the tool does, but did not detect the one in IAV segment 7.

As a first approach to assessing the strength of the segment 7 splice signals, we used a selection of splice site prediction tools. While the mRNA 2 SD site was predicted by the majority of the bioinformatic tools with some confidence, identification of the SA was less consistent and when found, often with low confidence (Table 1). Furthermore, none of the tools picked either the SA or mRNA 2 SD as the top candidate within the segment (data not shown). Of the two tools able to predict BP sites, HSF did not identify the site around A720 previously proposed to form a lariat structure (41) but instead identified the potential BP around A713 with reasonable confidence, while SVM-BP identified both candidates with low confidence but scored A713 more highly. SVM-BP also identified the same short 7 nucleotide PPT adjacent to the potential BP at A720 as previously proposed (41) with moderate confidence.

Since *in silico* analysis of segment 7 did not unambiguously identify which of the mRNA2 splice signals might be responsible for downregulating splicing efficiency, we next performed a systematic mutagenic analysis of each motif, starting from the 3'-end SA site. To avoid confounding effects from non-synonymous mutations in the M1 gene or overlapping segment-specific RNA packaging signals (54), we utilised a transfection-based minireplicon

system, using segment 7 as the reporter construct (35). To further mimic viral infection, we also included segment 8, given the potential effects of NS1 on segment 7 mRNA splicing, nuclear export and overall cellular translation (14, 35, 60-63).

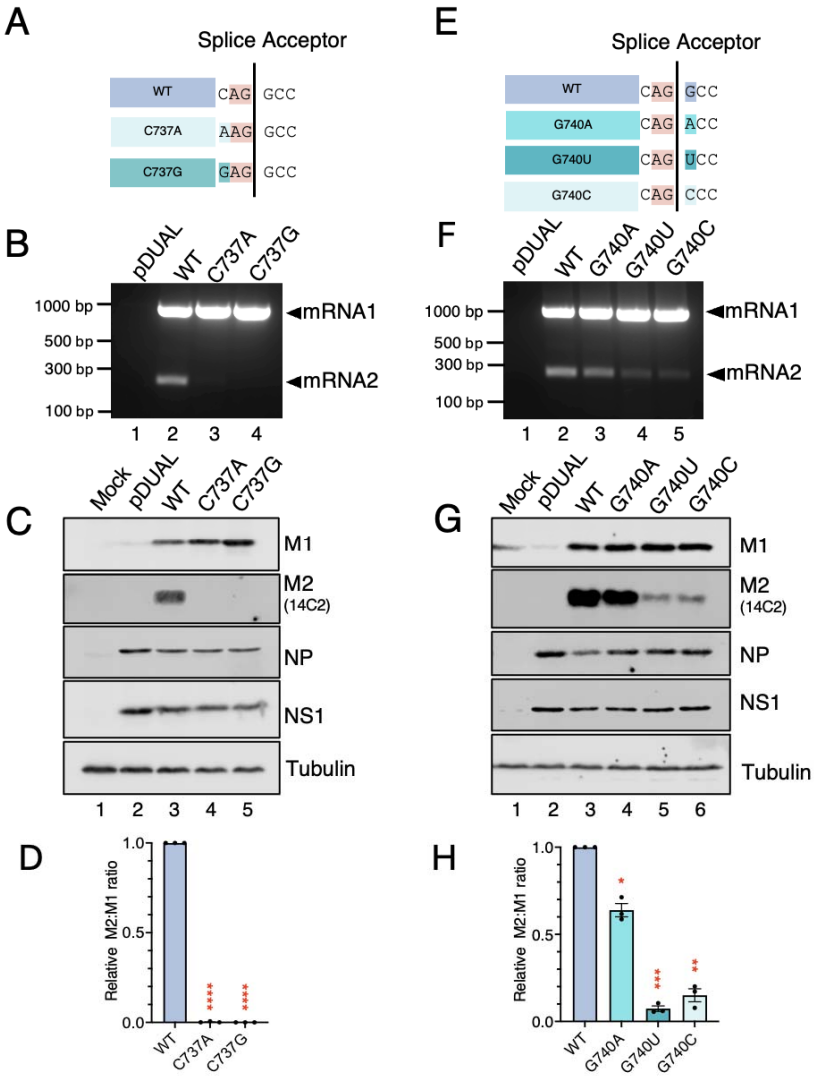

Figure 2. The segment 7 SA has an optimal sequence. (A and E) Design of mutations (blue shades) to change the nucleotide preceding or following the acceptor AG dinucleotide (orange). (B-D and F-H) HEK 293T cells were co-transfected (or mock transfected) with the viral polymerase and NP (3PNP) and NS segment plasmids along with the indicated WT or mutant segment 7 constructs (pDUAL indicates an empty vector control without segment 7) and harvested 48 h later. (B, F) Total RNA was extracted from cells and segment 7 mRNAs detected using RT-PCR followed by agarose gel electrophoresis. (C, G) Cells

were lysed in Laemmli buffer and protein expression was analysed by SDS-PAGE and western blotting. (D, H) Relative M2:M1 ratios were calculated by densitometry from three independent experiments. Statistical annotations are the result of an ordinary One-way ANOVA with Dunnett's post-test relative to WT. **** = P<0.0001; *** = P<0.001; ** = P<0.01; * = P<0.05; absence of significance symbols = P>0.05. Error bars indicate the standard error of the mean.

PR8 segment 7 has an optimal SA sequence

The core sequence of the PR8 segment 7 SA is an exact match to cellular consensus. However, to test for context-dependent effects, we examined the consequences of mutating the nucleotides either side of the essential AG motif, as not all cellular SA sites meet the extended consensus (Figure 1) (64, 65). Accordingly, two mutants were made in which the SA AG motif was preceded by either A or G at position 737 (Figure 2A). Uracil (U) was not tested as this would introduce a premature stop codon into the M1 gene. We then tested segment 7 RNA synthesis and protein expression from the mutated segments using the minireplicon system, measuring mRNA levels and relative M2:M1 ratios by RT-PCR and western blotting respectively. The WT segment produced a mix of unspliced mRNA1 and spliced mRNA2 (Figure 2B) as expected (35, 38), as well as appreciable quantities of both M1 and M2 polypeptides (Figure 2C). However, the C737A and C737G mutations strongly inhibited both mRNA splicing and M2 expression, reducing the latter to below the level of detection when replicate experiments were quantified (Figure 2D). Therefore, 737C is a key component of the segment 7 SA, consistent with its 100% conservation in segment 7 sequences (Figure 1) (66) and the likelihood of this nucleotide to cause under splicing in cellular mRNA (65).

In contrast to nucleotide 737, single nucleotide polymorphisms are seen at segment 7 nucleotide 740 after the AG, albeit with a clear preference for G and A over C or U (Figure 1).

To test whether these variations affect segment 7 splicing, we made constructs in which the WT 740G was replaced with A, U or C (Figure 2E) and monitored their splicing efficiency by RT-PCR and western blotting as before (Figure 2F-H). WT segment 7 WT exhibited the highest splicing efficiency by the qualitative RT-PCR assay (Figure 2F), while all 740 mutants had significantly lower splicing efficiency, as judged by quantitative western blots (Figures 2G, H). However, the G740A variant resulted in higher mRNA splicing efficiency higher than the other two mutants (Figure 2F-H). Thus, the G/A SNP at position 740 may contribute to regulation of segment 7 splicing in some strains of IAV. Altogether, none of the tested segment 7 mutations enhanced splicing efficiency, suggesting that IAV segment 7 SA is inherently strong, and any substitution in the nucleotides either side of the SA AG dinucleotide weaken splicing efficiency.

The segment 7 polypyrimidine tract is key to setting M1:M2 balance.

PPT sequences functions via recruiting spliceosomal U2AF, which then brings the U2 snRNP to the BP which is required for lariat formation (31, 32). The PPT is located upstream of the SA (29) and its strength is determined by proximity and continuity: a long PPT in close proximity to the SA is generally stronger than a short one or one further away from the SA (67). The PPT of segment 7 has been predicted to be nucleotides 721-727 of the pre-mRNA (underlined in Figure 1) (41), but this would be both short and relatively far from the SA. Alternatively, a more conventionally sited PPT ending two nucleotides upstream of the SA at position 735 would extend 5'-wards into the predicted PPT but is interrupted by 5 consecutive purines centred at position 730. To experimentally test these options, we first mutated pyrimidines within nucleotides 721 and 727 to either adenines or guanosines whilst avoiding the creation of cryptic SA (AG) motifs (Figure 3A, mutants 721-24A and 724-27G). As

before, pre-mRNA splicing in the minireplicon system was measured using RT-PCR and western blotting to detect M1 and M2 expression. Supporting the identification of the 721-727 region as a PPT, mRNA2 splicing and M2 protein production were decreased when pyrimidines were replaced by purines (Figure 3B lanes 3-4 and C lanes 4-5, Figure 3D for quantification). However, the two sets of mutations affected splicing differentially. While mRNA2 splicing was still detected for the 721-24A mutant, 724-27G almost totally abrogated splicing and subsequent M2 expression.

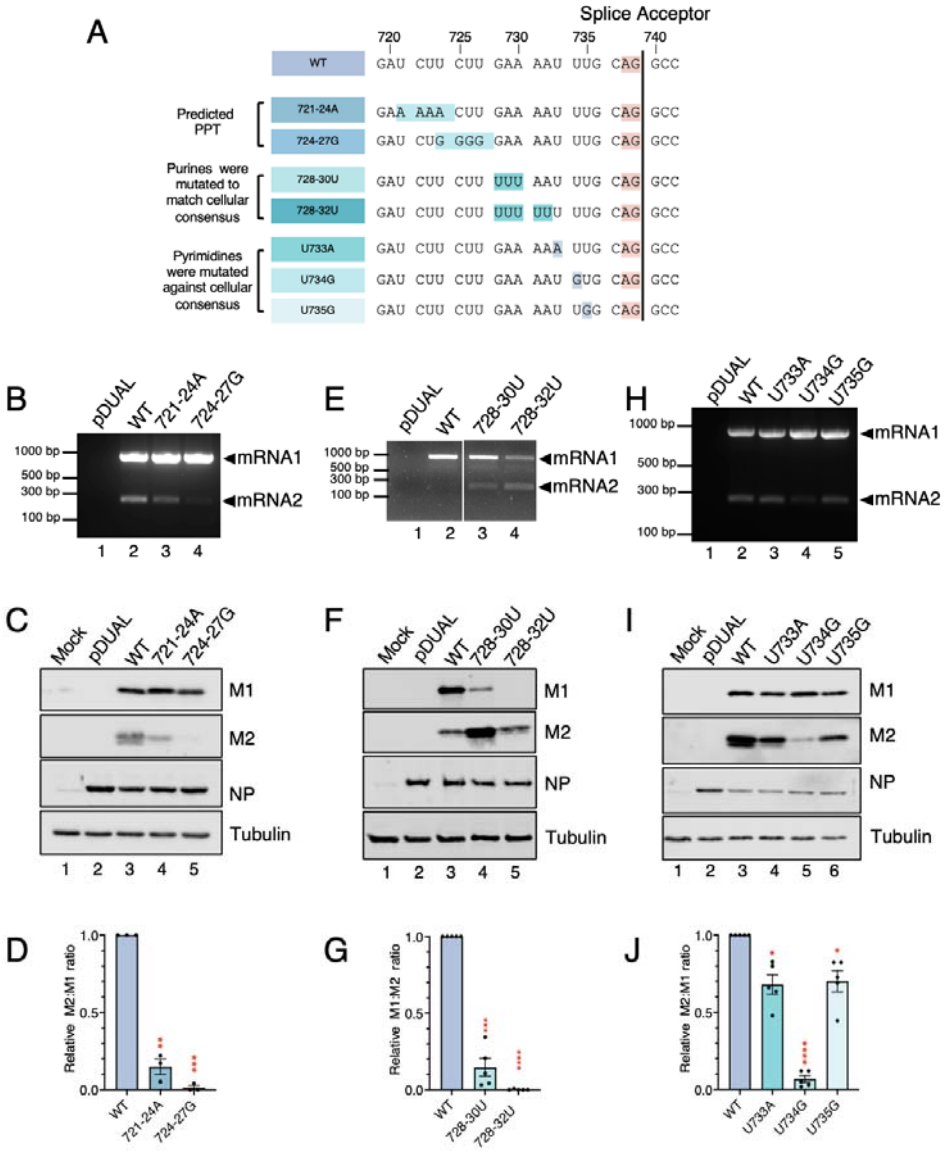

Figure 3. Identification of the segment 7 PPT. (A). Sequences of segment 7 PPT mutants with altered polypyrimidine tracts. Mutations are highlighted in blue shades and the segment 7 SA is highlighted in orange. (B-J) The RNP components, NS segment and the indicated segment 7 or pDUAL empty vector plasmids were co-transfected (or mock transfected) with the indicated segment 7 mutants in HEK 293T cells. (B, E, H) Segment 7 mRNA was detected by RT-PCR followed by agarose gel electrophoresis. (C, F, I) Protein expression was analysed by SDS-PAGE and western blotting. (D, G, J) Densitometric quantification of the relative M2:M1 ratios was analysed by an ordinary One-way ANOVA test with Dunnett's post-test relative to WT (N=3-5). **** = P<0.0001; *** = P<0.001; ** = P<0.01; * = P<0.05; absence of significance symbols = P>0.05. Error bars indicate the standard error of the mean.

However, while mutations within nucleotides 721-727 confirmed that the segment 7 PPT extends to position 721 upstream of the SA, they did not necessarily define its full extent. Therefore, we mutated regions downstream of the predicted PPT to either match (mutants 728-30U and 728-32U) or violate (U733A, U734G and U735G) the cellular PPT consensus. RT-PCR results from these constructs showed that mRNA1 levels decreased and mRNA2 levels increased with the introduction of uridines (Figure 3E, lanes 3-4), while removing uridines at positions 734 and 735 had the converse effect (Figure 3H, lanes 4 and 5), consistent with uridine content close to the SA site modulating splicing. Supporting this, all up and down mutations also caused significant alterations to the balance of M1 and M2 protein expression (Figures 3F, G, I and J; note that the y axis of G is reversed relative to panels D and J). However, while reduced mRNA1 levels in the 728-32U mutant correlated with a drastically lowered M1:M2 ratio, overall protein expression was low (Figure 3F, lane 5). This particular mutation introduced a 10 nucleotide long polyuridine stretch, which in other contexts has been shown to induce viral polymerase slippage (68-70), potentially providing an explanation for reduced levels of M1. In agreement, Sanger sequencing of PCR

products spanning the region showed a marked reduction in sequence fidelity at this region (data not shown).

Overall, these data indicate that segment 7 has a bi-partite PPT located between nucleotides 721-735, consisting of strong 5' and weaker 3' elements, with the latter crucial to preventing oversplicing of the primary transcript and lowered M1 expression.

Segment 7 has flexibly defined branch points

Splicing of cellular pre-mRNAs requires the formation of a lariat structure (29). The BP is an essential *cis*-acting element that interacts with the SD site to initiate lariat formation. In most cases, the BP is an adenine residue, as its 2'-OH group is necessary to attack the phosphate backbone of the SD sequence (29). Previous bioinformatic analysis predicted that nucleotide 720 might be the segment 7 BP (41). To test this, we mutated A720 to each of the other three nucleotides (Figure 4A; BP2). If nucleotide 720 were the sole BP, these mutations would be expected to destroy its function, block splicing and reduce M2 protein expression. Unexpectedly, no reduction in M2 expression was detected for any of the three mutants, which instead gave slightly increased (but not statistically significant) M2:M1 ratios relative to WT segment 7 (Figure 4B and C). Thus, either nucleotide 720 is not the segment 7 BP or an alternative BP can be utilised.

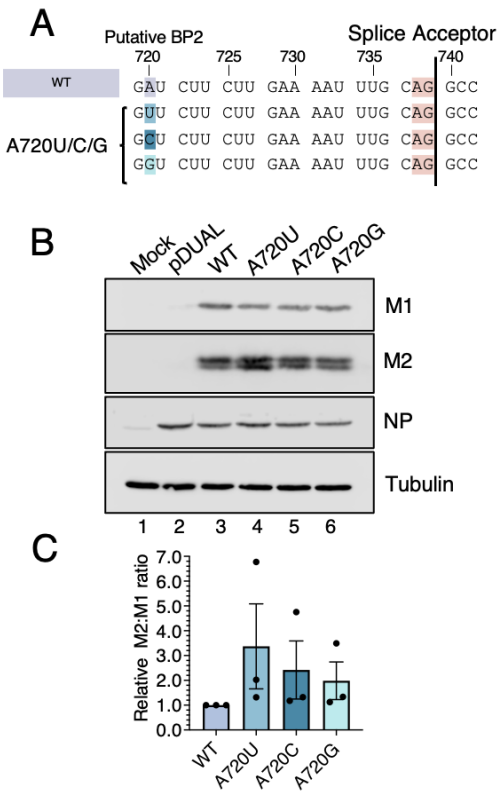

Figure 4. Mutational tests of putative BP2 at nucleotide 720. (A) Sequences of mutants with substitutions at the putative BP. Mutated sites are highlighted in blue and the segment 7 SA is highlighted in orange. (B) HEK 293T cells were co-transfected (or mock transfected) with the minireplicon components, NS segment and the indicated segment 7 or pDUAL empty vector plasmids. Cells were lysed and analysed by SDS-PAGE and western blotting as labelled. (C) Densitometric quantification of the relative M2: M1 ratio was analysed by an ordinary One-way ANOVA test with Dunnett's post-test relative to WT (N=3). Absence of significance symbols = P>0.05. Error bars indicate the standard error of the mean.

As noted above (Figure1 and Table 1), another potential BP in segment 7 is predicted at nucleotides 713, in addition to the one at position 720. To systematically test if these adenosines can act redundantly as the segment 7 BP, we mutated them both to Gs, either as

single or double mutants (mutants A731G, A720G and A713G+A720G) (Figure 5A). As before (Figure 4), mutation of A720 did not reduce mRNA2 or M2 expression. However, decreases in mRNA2 levels and M2 expression were seen in the A713G and A713G+A720G mutants (Figure 5B, lanes 3 and 5 and Figure 5C, lanes 4 and 6; quantification in 5D), although without completely blocking splicing. This result suggests that while A713 is important for splicing, it is not the only potential BP residue. Instead, the segment 7 BP could be loosely defined (71), and nearby adenosines could be used in the absence of canonical BPs (72). To test this, we mutated two additional adenosines at nucleotides 714 and 715 (A713-5G). This mutant showed a more pronounced drop in mRNA2 and M2 levels, but low levels of splicing still occurred (Figures 5B, lane 6; 5C lane 7, 5D). However, further addition of the A720G mutation (A713-5G+A720G) led to barely detectable mRNA2 levels and undetectable M2 expression (Figure 5B, lane 7, Figure 5C lane 8, Figure 5D). These results support the hypothesis that segment 7 can use more than one BP adenosine, at positions 713-715 and 720.

A previous study identified serine and arginine-rich splicing factor 5 (SRSF5) to be an mRNA2 splicing factor that interacts with nucleotides 707-715 (73), overlapping our putative BPs. To distinguish whether the absence of mRNA2/M2 in the A713-5G+A720G mutant was caused by the loss of BP function or potential lack of SRSF5 binding, we further tested the mutations in the context of mRNA4 splicing, which does not depend on SRSF5 (73). However, given that WT PR8 only produces low levels of mRNA4, we altered the segment splicing pattern to improve M42 expression. To achieve this, the mRNA4 SD was improved through the introduction of a U148A and the mRNA2 SD was destroyed with a G52C change (22). As expected, single mutation of U148A upregulated mRNA4 at the expense of mRNA2, without abolishing its production (Figure 5E, lane 4), while G52C severely reduced mRNA2 levels (Figure 5E, lane 3). The combination of the two mutations reduced mRNA2 below the

level of detection (Figure 5E, lane 5). We then added the set of A713-5G-A720 potential BP adenosine mutations on these two backgrounds. This caused a similar reduction in mRNA4 levels to that seen for mRNA2 (compare lane 6 with 7 and 8). Next, we used two anti-M2 antibodies with differing specificities to differentiate accumulation of the M2 and M42 proteins: the 14C2 antibody recognises residues 4 to 17 of the PR8 M2 protein but does not bind M42 (22, 74), while a polyclonal antibody (PA5-32233) that was raised against amino acids 11-97 of M2 recognises both M2 and M42 but is unable to distinguish them. M1 levels were relatively consistent across all samples transfected with segment 7 plasmids, while M2/M42 levels showed greater variation. The G52C mutant did not produce detectable amounts of M2 and its M42 protein expression was also below the detection limit (Figure 5F, lane 4). For the U148A mutant, 14C2 reactivity was weaker than the WT, whereas PA5-32233 reactivity was comparable to WT, indicating that U148A produced both M2 and M42 (Figure 5F, lane 5). The double mutant G52C-U148A produced only M42, as evidenced by reactivity with PA5-32233 but not with 14C2 (Figure 5F, lane 6). Consistent with the RT-PCR data, once 713-715 and 720 adenosines were mutated, both M2 and M42 production were dramatically reduced (Figure 5F, lanes 7-9; quantification in 5G). Given that SRSF5 does not affect mRNA4 splicing and M42 expression (73), the much decreased mRNA4 and M42 production seen here suggests that adenosines 713-715 and 720 play a distinct role in segment 7 splicing beyond serving as an SRSF5 binding motif. Overall, these data suggest that segment 7 has redundancy in BP function to mediate mRNA2 and mRNA4 splicing.

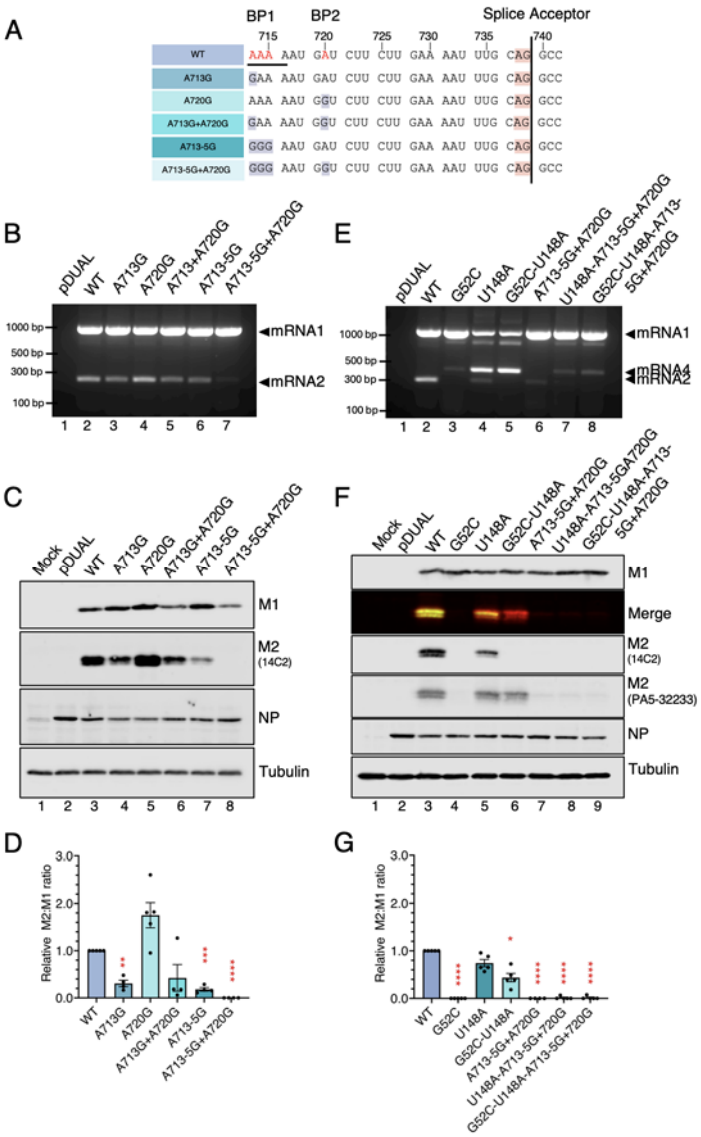

Figure 5. Mutational identification of the segment 7 BP. (A) Mutational strategy for identifying BP adenosines in segment 7. Adenosines that were predicted as potential BPs are shown in red on the WT sequence, mutations are highlighted in purple, while the SA is highlighted in orange. The region of the SRSF5 binding motif overlapping with BP mutations is underlined. (B-G) HEK 293T cells were co-transfected (or mock transfected) with the minireplicon components, NS segment and the indicated segment 7 or pDUAL empty vector plasmids, harvested at 48h and analysed by (B, E) RT-PCR and agarose gel electrophoresis and (C, F) western blotting as labelled. (D, G) M2:M1 ratios were quantified by densitometry of replicate experiments and normalised to the value of the WT construct. The relative M2: M1 ratios were analysed by an ordinary One-way ANOVA test with Dunnett's post test relative to WT

(N=4-5). **** = P<0.0001; *** = P<0.001; ** = P<0.01; * = P<0.05; absence of significance symbols = P>0.05. Error bars indicate the standard error of the mean.

The strength of the segment 7 SD is set by nucleotides C51 and C55

The SD motif of cellular pre-mRNAs is recognised by base-pairing between the U1 snRNP and 9 nucleotides either side of the mRNA splice site (Figure 6A) (28). The IAV segment 7 mRNA2 SD sequences are highly conserved (except for nucleotide 55) and a close match to cellular consensus. The two exceptions are nucleotides 51 and 55 (Figure 6B), although the presence of a pseudo-uridine in U1 snRNA may allow some flexibility in base pairing for the latter residue.

To systematically investigate how the mRNA2 SD sequence might regulate segment 7 splicing, we created a panel of all possible point mutations in the mRNA2 SD region and analysed their behaviour in the minireplicon assay as before. RT-PCR analysis of segment 7 mRNAs indicated that altering the three nucleotides in the exon side of the motif had only slight effects on splicing levels, although mutation C51G made splicing slightly stronger (mRNA2:mRNA1 ratio 2.3 ± 0.3-fold (n=3) higher than WT) (Figure 6C, lanes 11). However, most substitutions in the intron side of the SD motif influenced splicing (Figures 6D and E). All variations of G52, U53, G56 and U57 inhibited mRNA2 splicing, mostly drastically (Figure 6D, lanes 3-8 and Figure 6E, lanes 6-11). Defective recognition of the mRNA2 SD was also associated with upregulation of mRNA4 levels, presumably due to competition between these two SDs for U1 snRNP interaction (22). Conversely, C55, the only intronic position with any variability in segment 7 (Figure 6B), was more tolerant of mutation. Consistent with the IAV consensus, C55U showed a similar level of splicing to WT, while mutating C to a purine (especially A) made splicing stronger (mRNA2:mRNA1 ratio 3.1 ± 0.1-fold higher than WT, n = 3). In addition, we used western blotting to measure viral

protein expression. Since some of the introduced point mutations are non-synonymous in M1/M2 and amino acid substitutions in the M2 ectodomain interrupt the 14C2 anti-M2 antibody epitope, the polyclonal anti-M2 antibody (PA5-32233) was used. NP and NS1 levels were similar between all samples, indicating equivalent levels of transfection (Figures 6F-H). Consistent with the RNA analyses, point mutation C51G reduced M1 accumulation and gave higher levels of M2, suggesting mRNA1 oversplicing (Figure 6F; compare lanes 4 and 13). The C55A and C55G substitutions formed premature stop codons in M1 so predictably, no M1 was detected in these two samples (Figure 6H, lanes 5 and 7). Even so, C55A showed increased M2 levels, consistent with the increased mRNA2 splicing detected by RT-PCR. Also consistent with the RT-PCR data, M2 antibody reactivity was absent (or very much reduced; U57 series) from all mutants having mRNA2 splicing defects (Figure 6G, lanes 5-12 and 6H, lanes 8-13). However, even though defective mRNA2 SD sequences upregulated mRNA4 SD recognition, M42 levels did not reach the detection limit in our experiments.

Thus, combining the RT-PCR and western blotting results, all point mutations of nucleotides 52, 53, 54, 56 and 57 inhibited mRNA2 splicing. Only two mutations, C51G and C55A, increased segment 7 splicing. Thus, the mRNA2 SD is strong, but not entirely optimised to support mRNA2 SD recognition and splicing.

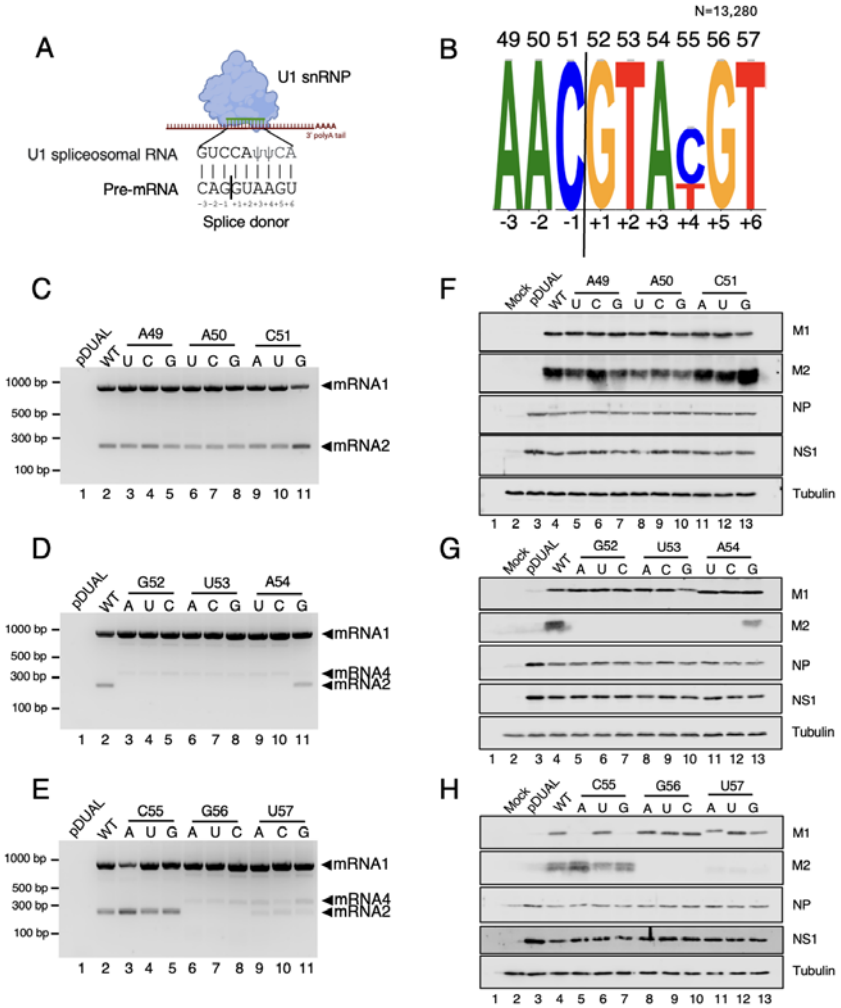

Figure 6. Mutational analysis of the mRNA2 SD sequence. (A) Cartoon diagram of base-pairing between U1 snRNP and mRNA SD sequences. Ψ represents pseudouridine, an isomer of uridine found in U1 snRNA. (B) Sequence logo indicating conservation of the segment 7 mRNA SD site, generated from 13,280 IAV segment 7 sequences. (C-H) The mRNA2 SD was systematically mutated and individual constructs were tested by minireplicon system. HEK 293T cells were co-transfected (or mock transfected) with viral polymerase and NP (3PNP), NS segment and the indicated segment 7 or pDUAL empty vector plasmids. Lysates were harvested 48h post transfection and (C-E) segment 7 mRNA detected by RT-PCR followed by agarose gel electrophoresis. (F-H). Viral protein expression levels were analysed using western blotting.

Splicing regulation in viable viruses

So far, we analysed segment 7 mRNA splicing using a minireplicon system to avoid any confounding effects of splice site mutations on the primary amino-acid sequence of M1 and segment-specific packaging signals. However, the kinetics and regulation of segment splicing in the minireplicon system may not fully reflect that of the authentic virus. Therefore, to test how the *cis*-acting splicing elements regulate mRNA production in the context of viral infection, as well as to explore the evolutionary conflicts between the multiple functional elements in segment 7, we attempted to rescue a panel of viruses containing mutations that altered the balance of mRNA1 and mRNA2 production in the minireplicon system. For the PPT, we redesigned the mutations to break homopolymeric adenosine runs with guanosines (Figure 7A; 728C-30U, 728C+730U, 731-32Y and 728C+730-32Y), to avoid the potential for transcriptional slippage by the viral polymerase (70).

Viruses were rescued by transfecting bidirectional reverse genetics plasmids into HEK 293T cells, amplified by one passage in MDCK cells, and plaque titred. WT PR8 grew to an average titre of 6×10^{8}PFU/ml and formed large plaques (Figures 7B, C). No mutants with alterations to the SD motif or the preceding nucleotide of the SA site could be successfully rescued. Two of three mutants with altered first nucleotides downstream of the SA in the exon grew to WT-like levels, albeit with small-plaque phenotypes, but G740C could also not be rescued. Most PPT mutants rescued successfully with only 728-30U giving similar titres to WT, albeit forming significantly smaller plaques. 728C+ 730-32Y could be rescued but replicated to 10-fold lower titres than WT, while 734-36R was apparently non-viable. BP2 mutations at position 720 rescued readily and replicated to titres approaching that of WT virus, with only moderate small plaque phenotypes. The BP1 mutation A713G caused a far more severe growth defect. Overall, more than half the splice mutations gave viable viruses

with only moderate fitness penalties, indicating a degree of evolutionary flexibility in the 3' splice site sequence motifs.

Most of the mutations in the splicing motif are non-synonymous in M1 protein, which could potentially influence both splicing signals and M1 function. To assess the potential effects of these mutations on M1 functions we mapped our mutations onto the output (75) of a deep mutational scan of the PR8 M1 protein (Figure 7D). Some of the sites altered in our study exhibited a high mutational tolerance, for example, D232 (A720U, A720C and A720G), E235 (728-30U and 731-32Y), suggesting a lower likelihood of impact on M1 function. In contrast, the lower diversity of amino acids recovered at positions 9 - 11 of M1 suggests greater functional conservation and could be one of the reasons why the SD mutants tested were lethal. This defect could be caused by impaired splicing, compromised M1 function, or a combination of both.

Next, we investigated segment 7 mRNA splicing levels in cells infected with the set of WT and mutant viruses that grew to sufficient titres to permit high multiplicity infection. NP mRNA was analysed as an infection control, and this confirmed that all infections had proceeded successfully (Figure 7E). In cells infected with WT PR8, a mixture of unspliced mRNA1 and spliced mRNA2 was detected (Figure 7E, lane 2). Mutations at the nucleotide immediately following the AG (especially G740U) resulted in increased mRNA1 levels and reduced mRNA2 production, confirming that this position serves as an essential *cis*-acting element for regulating splicing (Figure 7E, lanes 3 and 4). In contrast, mutant viruses containing additional pyrimidines in the PPT enhanced splicing efficiency, thereby shifting the mRNA1 and mRNA2 production ratio in viable viruses. The A720G BP mutant had relatively lower levels of NP mRNA, suggesting impaired infectivity. However, mRNA2 was still detectable in all the A720 mutant viruses. This result was consistent with the previous

observation in the non-infectious mini-replicon system, indicating that nucleotide 720 is either not the BP or not the only BP for segment 7 splicing. Thus overall, the RT-PCR results for segment 7 splicing were consistent with our previous minireplicon assay results.

Next, we analysed segment 7 protein expression from the mutant viruses by western blotting. NP and tubulin levels were comparable in most samples, suggesting successful infection and recovery of cellular protein. However, M2 levels were far more variable. The G740A and G740U mutations flanking the AG acceptor dinucleotide gave very low and undetectable amounts of M2, (Figure 7F, lanes 3 and 4 respectively). Quantification of replicate experiments confirmed that the M2:M1 ratio of these two mutants was significantly lower than that of WT, indicating M2 protein expression defects. Conversely, the 728-30U and 731-32Y PPT up-mutations led to abundant synthesis of M2 proteins (Figure 7F, lanes 5-6). The BP2 mutations at position 720 gave conflicting phenotypes. A720U and A720C matched the minireplicon and viral mRNA results, producing abundant quantities of M2 and a small (but not statistically significant) increase in M2:M1 ratios. However, in the context of virus, A720G drastically underproduced M2, despite only a slight decrease in mRNA2 levels. Nevertheless, M2 expression in all BP mutants confirm that A720 is not the sole segment 7 BP.

Overall, the data from viral infections were broadly consistent with our previous minireplicon results and thus support the proposed mechanism by which *cis*-acting elements in segment 7 are the primary regulator of mRNA splicing. They also reveal that the virus has a surprising degree of tolerance for altered M1 to M2 ratios.

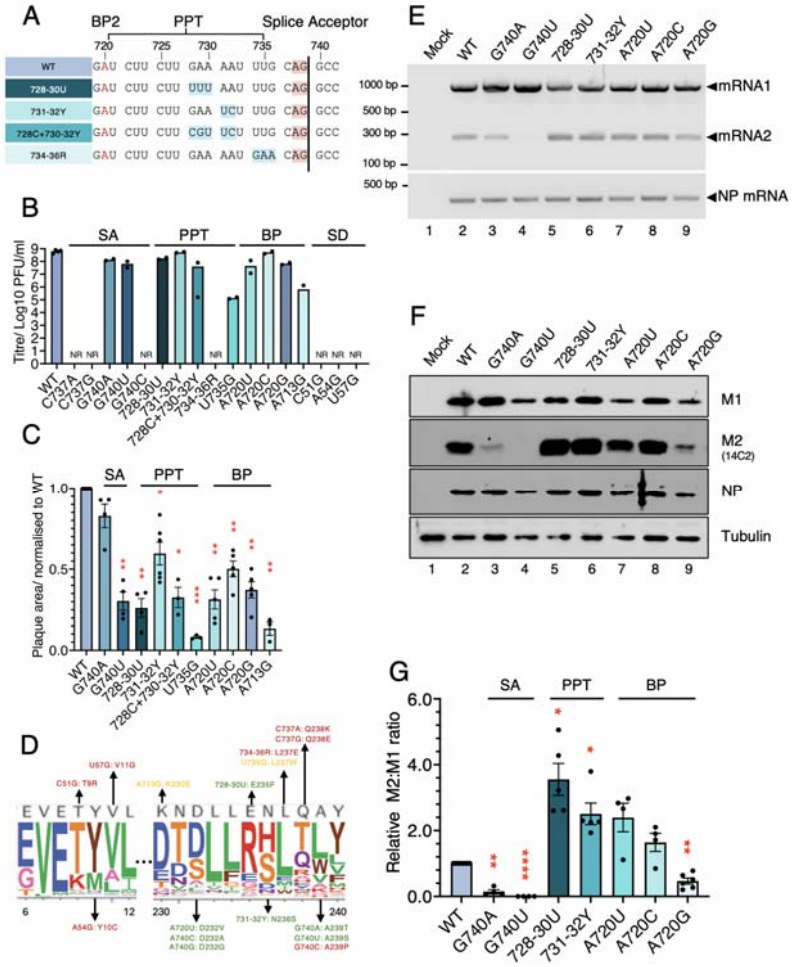

Figure 7. Virus phenotypes following mutation of *cis*-acting segment 7 splice elements. (A) Design of a new panel of mutants with mutated polypyrimidine tracts. Mutations are highlighted in blue. The canonical splice acceptor site is shown in orange and the positions of BP2 and PPT are labeled. (B) P1 stock titres of WT and mutant viruses. Data represent the mean of at least two independent rescues. (C) Quantification of the plaque sizes of PR8 WT and mutant viruses. Data are the mean ± SEM of 5-20 plaques normalised to the average WT value from at least three independent experiments. (D) Logo plot representing the amino acid preferences at each M1 residue site (75). Mutations tested here are labelled, and amino acid substitutions are shown. Mutants that could not be rescued are shown in red; mutants exhibiting titres less than 10% of WT are coloured in amber; mutants which grew to near WT levels (sufficient to allow high multiplicity synchronous infections) are coloured in green. (E) The effects of splice element mutations on viral mRNA splicing. MDCK cells were infected (or mock-infected) with the

indicated viruses at an MOI of 5. Total RNA was isolated at 6 h p.i and the indicated mRNAs were detected by RT-PCR. (F) MDCK cells were infected (or mock-infected) with the viruses at an MOI of 5, and lysated at 8 h p.i. The viral protein expressions were analysed by SDS-PAGE and western blotting. (G) Quantification of the relative M2: M1 ratio was conducted using an ordinary One-way ANOVA test with Dunnett's post test relative to WT (N=4-6). **** = P<0.0001; *** = P<0.001; ** = P<0.01; * = P<0.05; absence of significance symbols = P>0.05. Error bars indicate the standard error of the mean. Panel D was adapted with permission from Hom et al., Journal of Virology 2019;93:e00161-19 (CC BY 4.0 https://creativecommons.org/licenses/by/4.0/).

Discussion

IAV uses alternative splicing of segment 7 to encode two proteins, M1 and M2, and impaired production of either of these two proteins can reduce virus fitness (22, 76-79). The regulation of segment 7 alternative splicing is complex, and previous studies have reported various control mechanisms, involving *cis*-acting sequences in segment 7 and *trans*-acting protein factors, both host and viral. For example, a polypyrimidine and polycytosine tract downstream of the mRNA2 SD has been shown to influence splicing by binding to the NS1-BP and hnRNP K proteins (34). Alternative splicing factor 1/splicing factor 2 (ASF/SF2) which binds to the downstream purine-rich stretch of SA is also involved in mRNA2 splicing as a *trans*-acting element (36). Potential viral *trans*-acting factors include the viral polymerase binding to the cRNA 5'-promoter element downstream of the 5'-cap-snatched leader sequence to block usage of the mRNA3 SD site, itself proposed to regulate usage of the mRNA2 SD (37). NS1 also potentially modulates splicing through direct or indirect interactions with mRNA1 and mRNA2 as well as by promoting nuclear export of unspliced mRNA1 (35, 80-82). There may be elements of virus strain-specificity to some of these mechanisms, given (for instance) the variable amounts of mRNA3 production and

disagreements over whether or not M2 is essential for virus replication (22, 78). Analyses of the role of NS1 in controlling viral mRNA splicing have also given variable results (38).

Here, we describe how suboptimal splice signals in IAV segment 7 maintain the desired ratio of unspliced and spliced transcripts and consequently balanced production of M1 and M2 proteins (Figure 8). A strong SA and redundant BP adenosines promote efficient mRNA2 splicing and M2 protein production. Negative sequence elements in the SD and PPT prevent oversplicing to preserve sufficient essential viral protein M1 production from unspliced mRNA.

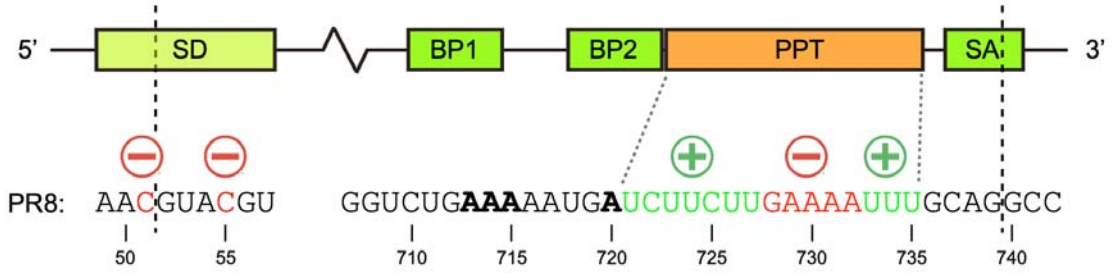

Figure 8. Diagrammatic summary of how IAV segment 7 uses suboptimal splicing signals to maintain the balance of mRNA1/M1 and mRNA2/M2 production. Segment 7 has a strong splice acceptor (SA), redundant and strong branch points (BPs; potential lariat-forming adenosines indicated in bold) and a medium-strong splice donor (SD) which allow the virus to produce sufficient M2 protein. A purine-rich stretch in a medium strength polypyrimidine tract (PPT) prevents oversplicing, thereby preserving levels of unspliced mRNA1 required for the synthesis of the essential M1 protein.

Thus, although multiple factors can regulate segment 7 mRNA splicing, we propose that the four required *cis*-acting elements, SD, BP, PPT, SA are the primary determinants that

regulate viral mRNA splicing. This model is consistent with the high conservation of the segment 7 splice motifs (Figure 1B) and would maintain a basal regulation of M1/M2 production across the frequent reassortment events seen in avian and other influenza A viruses that could pair the segment with variants of the viral polymerase and/or NS1 with altered splicing regulatory activity. Given the evolutionary conservation of the basal splicing machinery across birds and mammals (83, 84), segment-autonomous regulation may also help underpin a broad host range. Further "fine-tuning" of splicing ratios could then evolve for the particular gene constellation in an individual virus lineage or host.

Achieving the correct balance of M1 and M2 exerts a selection pressure on the virus to evolve suboptimal splicing signals. Further sequence constraints arise from maintaining function of the translation products encoded by the mRNAs, as well as any other *cis*-acting functions of the segment. All splice site elements considered in this study overlap with the M1 open reading frame (ORF) and some of them overlap the M2 ORF. Consequently, most mutations that alter splicing efficiency are non-synonymous in M1 and sometimes in M2. The mRNA2 SD sequence is also involved in the segment-specific packaging signal (54). Mutations in this region therefore might not only change splicing efficiency (27) but also affect virus genome packaging, or even M1/M2 translation via codon bias effects. Consistent with a severe constraint on the segment sequence, all our attempts to rescue three mRNA2 SD mutants failed. A previous study also reported failure to rescue mRNA2 SD mutants, finding that virus could only be generated successfully when the mutations within mRNA2 SD motif (G52 and U53, M1 codons 9 and 10) were synonymous (16). Other mutants which also failed to produce mRNA2 but which also altered codons 9 or 10 of M1 were non-viable. Therefore, either M1 sequence and/or packaging signal sequences are critical in this region, hampering modifications to the splice site.

In contrast to the relative mutational inflexibility of the SD sequence, the purine-rich stretch (728-733 of segment 7) in PPT tract was much more tolerant of changes. When we introduced pyrimidines, which are non-synonymous in M1, to this region, viable viruses with only limited fitness defects were produced; WT-level titres but small-plaque phenotypes, indicating mutational flexibility in this region. This is consistent with the output of the deep mutational scan of PR8 M1 protein where M1 E235 and N236 sites exhibited a high mutational tolerance for amino acid substitutions (75) and (Figure 7D). Furthermore, this region does not overlap with any currently defined packaging signals in segment 7 (85, 86). Thus, unlike the SD region, which plays multiple essential roles (packaging signal, splice site and encoding M1 and M2) during the viral life cycle, the PPT region has "only" two defined functions: M1 protein coding and acting as a PPT. The combination of relative functional simplicity and high mutational tolerance suggests that the virus can more easily regulate splicing by changing the PPT sequence during evolution. This might explain why the PPT is the most important down-regulator of splicing as it allows more flexibility without severe impacts on viral fitness.

Although the majority of mutations in this study were introduced into the M2 intron, expression of M2 was directly affected by changes in splicing efficiency. In most situations, M2 is considered essential for normal virus replication (22, 78). However, our data show that the virus can tolerate a low M2 expression level and still replicate to high titres. The G740U mutant exhibited a severe splicing defect and failed to produce detectable amounts of mRNA2/M2 during high multiplicity synchronous infection. Despite this, it only exhibited modest growth defects: a slightly reduced titre compared to WT PR8 and a small-plaque phenotype. These data suggest that the PR8 virus only needs a very low level of M2 when replicating in tissue cultured cells. This finding aligns with previous studies (16, 22, 54, 78, 79), showing a variety of IAV strains can undergo multiple cycles of replication despite

lacking M2 ion channel activity or overall M2 expression. However, the degree of replication attenuation seen varies by over 1000-fold, depending on strain and mutational strategy, raising the possibility that the requirement for M2 may vary across different strains.

To avoid confounding effects from mutations affecting M1 or M2 protein function and/or packaging signals, we based most of our study around a non-infectious minireplicon system, which may not completely recapitulate the temporal aspects of virus infection (15, 76, 87). Nevertheless, when we compared our results from minireplicon and viable virus infection, the majority of mutants behaved similarly. The exception was the A720G mutation, which showed different splicing patterns under the two conditions. In the minireplicon system, A720G produced slightly more mRNA2 and an elevated (though not statistically significant) M2: M1 ratio (Figures 4B and 5B). In contrast, during single-cycle infection, A720G produced seemingly normal amounts of mRNA2 but had a significantly reduced M2:M1 ratio (Figure 7E). Potentially, this could reflect cell type differences, as the minireplicon system used HEK 293T cells while infection was carried out in MDCK cells (13, 15, 76). Further experiments are required to test this hypothesis.

Acknowledgements

We thank Professor Chris Smith and Dr Colin Sharp for helpful discussion.

Funding

RMP, ERG and PD acknowledge Institute Strategic Programme grant support from the UK Biotechnology and Biological Sciences Research Council (grant no. BBS/E/RL/230002C).

Bibliography

1. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza. Nat Rev Dis Primers. 2018;4(1):3.
2. Pinto RM, Lycett S, Gaunt E, Digard P. Accessory gene products of influenza A virus. Cold Spring Harbor perspectives in medicine. 2021;11(12):a038380.
3. Chen BJ, Leser GP, Jackson D, Lamb RA. The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. Journal of virology. 2008;82(20):10059-70.
4. Wachsmuth-Melm M, Peterl S, O'Riain A, Makroczyová J, Fischer K, Krischuns T, et al. Visualizing influenza A virus assembly by in situ cryo-electron tomography. Nature Communications. 2025;16(1):9394.
5. Bourmakina SV, García-Sastre A. Reverse genetics studies on the filamentous morphology of influenza A virus. Journal of General Virology. 2003;84(3):517-27.
6. Elleman C, Barclay W. The M1 matrix protein controls the filamentous phenotype of influenza A virus. Virology. 2004;321(1):144-53.
7. Martin K, Heleniust A. Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. Cell. 1991;67(1):117-30.
8. Pinto LH, Holsinger LJ, Lamb RA. Influenza virus M2 protein has ion channel activity. cell. 1992;69(3):517-28.
9. Hay A, Wolstenholme A, Skehel J, Smith MH. The molecular basis of the specific anti□influenza action of amantadine. The EMBO journal. 1985;4(11):3021-4.
10. Rossman JS, Jing X, Leser GP, Lamb RA. Influenza virus M2 protein mediates ESCRT-independent membrane scission. Cell. 2010;142(6):902-13.
11. Gannagé M, Dormann D, Albrecht R, Dengjel J, Torossi T, Rämer PC, et al. Matrix protein 2 of influenza A virus blocks autophagosome fusion with lysosomes. Cell host & microbe. 2009;6(4):367-80.
12. Hutchinson EC, Charles PD, Hester SS, Thomas B, Trudgian D, Martínez-Alonso M, et al. Conserved and host-specific features of influenza virion architecture. Nature Communications. 2014;5(1):4816.
13. Calderon BM, Danzy S, Delima GK, Jacobs NT, Ganti K, Hockman MR, et al. Dysregulation of M segment gene expression contributes to influenza A virus host restriction. PLoS Pathog. 2019;15(8):e1007892.
14. Mor A, White A, Zhang K, Thompson M, Esparza M, Muñoz-Moreno R, et al. Influenza virus mRNA trafficking through host nuclear speckles. Nature microbiology. 2016;1(7):1-13.
15. Valcárcel J, Portela A, Ortín J. Regulated M1 mRNA splicing in influenza virus-infected cells. Journal of general virology. 1991;72(6):1301-8.
16. Cheung TKW, Guan Y, Ng SSF, Chen H, Wong CHK, Peiris JSM, et al. Generation of recombinant influenza A virus without M2 ion-channel protein by introduction of a point mutation at the 5' end of the viral intron. J Gen Virol. 2005;86(Pt 5):1447-54.
17. Inglis SC, Brown CM. Spliced and unspliced RNAs encoded by virion RNA segment 7 of influenza virus. Nucleic acids research. 1981;9(12):2727-40.
18. Lamb RA, Lai C-J, Choppin PW. Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. Proceedings of the National Academy of Sciences. 1981;78(7):4170-4.
19. Shih S-R, Suen P-C, Chen Y-S, Chang S-C. A novel spliced transcript of influenza A/WSN/33 virus. Virus genes. 1998;17:179-83.

20. Bogdanow B, Wang X, Eichelbaum K, Sadewasser A, Husic I, Paki K, et al. The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. Nature Communications. 2019;10(1):5518.
21. Jackson D, Lamb RA. The influenza A virus spliced messenger RNA M mRNA3 is not required for viral replication in tissue culture. J Gen Virol. 2008;89(Pt 12):3097-101.
22. Wise HM, Hutchinson EC, Jagger BW, Stuart AD, Kang ZH, Robb N, et al. Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. PLoS Pathog. 2012;8(11):e1002998.
23. Wilkinson ME, Fica SM, Galej WP, Norman CM, Newman AJ, Nagai K. Postcatalytic spliceosome structure reveals mechanism of 3′–splice site selection. Science. 2017;358(6368):1283-8.
24. Hall SL, Padgett RA. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. Elsevier; 1994. p. 357-65.
25. Dietrich RC, Incorvaia R, Padgett RA. Terminal intron dinucleotide sequences do not distinguish between U2-and U12-dependent introns. Molecular cell. 1997;1(1):151-60.
26. Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. Are snRNPs involved in splicing? Nature. 1980;283(5743):220-4.
27. Lear A, Eperon L, Wheatley I, Eperon I. Hierarchy for 5′ splice site preference determined in vivo. Journal of molecular biology. 1990;211(1):103-15.
28. Wong MS, Kinney JB, Krainer AR. Quantitative Activity Profile and Context Dependence of All Human 5' Splice Sites. Mol Cell. 2018;71(6):1012-26 e3.
29. Brow DA. Allosteric cascade of spliceosome activation. Annual review of genetics. 2002;36(1):333-60.
30. Sperling J, Azubel M, Sperling R. Structure and function of the Pre-mRNA splicing machine. Structure. 2008;16(11):1605-15.
31. Zamore PD, Green MR. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proceedings of the National Academy of Sciences. 1989;86(23):9243-7.
32. Zamore PD, Patton JG, Green MR. Cloning and domain structure of the mammalian splicing factor U2AF. Nature. 1992;355(6361):609-14.
33. Tsai P-L, Chiou N-T, Kuss S, García-Sastre A, Lynch KW, Fontoura BM. Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. PLoS pathogens. 2013;9(6):e1003460.
34. Thompson MG, Muñoz-Moreno R, Bhat P, Roytenberg R, Lindberg J, Gazzara MR, et al. Co-regulatory activity of hnRNP K and NS1-BP in influenza and human mRNA splicing. Nature communications. 2018;9(1):2407.
35. Pereira CF, Read EKC, Wise HM, Amorim MJ, Digard P. Influenza A Virus NS1 Protein Promotes Efficient Nuclear Export of Unspliced Viral M1 mRNA. J Virol. 2017;91(15).
36. Shih SR, Krug RM. 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. 1996;15(19):5415-27.
37. Shih S-R, Nemeroff ME, Krug RM. The choice of alternative 5'splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex. Proceedings of the National Academy of Sciences. 1995;92(14):6324-8.
38. Robb NC, Fodor E. The accumulation of influenza A virus segment 7 spliced mRNAs is regulated by the NS1 protein. Journal of general virology. 2012;93(1):113-8.
39. Lamb RA, Lai C-J. Spliced and unspliced messenger RNAs synthesized from cloned influenza virus M DNA in an SV40 vector: expression of the influenza virus membrane protein (M1). Virology. 1982;123(2):237-56.

40. Valcárcel J, Fortes P, Ortín J. Splicing of influenza virus matrix protein mRNA expressed from a simian virus 40 recombinant. Journal of general virology. 1993;74(7):1317-26.
41. Moss WN, Dela-Moss LI, Priore SF, Turner DH. The influenza A segment 7 mRNA 3' splice site pseudoknot/hairpin family. RNA Biol. 2012;9(11):1305-10.
42. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome research. 2004;14(6):1188-90.
43. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic acids research. 1990;18(20):6097-100.
44. Kadri NK, Mapel XM, Pausch H. The intronic branch point sequence is under strong evolutionary constraint in the bovine and human genome. Communications biology. 2021;4(1):1206.
45. Saxonov S, Daizadeh I, Fedorov A, Gilbert W. EID: the Exon–Intron Database—an exhaustive database of protein-coding intron-containing genes. Nucleic acids research. 2000;28(1):185-90.
46. Wang M, Marín A. Characterization and prediction of alternative splice sites. Gene. 2006;366(2):219-27.
47. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. Journal of molecular biology. 1997;268(1):78-94.
48. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research. 2009;37(9):e67-e.
49. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouzé P, Brunak S. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic acids research. 1996;24(17):3439-52.
50. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. Journal of molecular biology. 1991;220(1):49-65.
51. Reese MG, Eeckman FH, Kulp D, Haussler D, editors. Improved splice site detection in Genie. Proceedings of the first annual international conference on Computational molecular biology; 1997.
52. Scalzitti N, Kress A, Orhand R, Weber T, Moulinier L, Jeannin-Girardon A, et al. Spliceator: multi-species splice site prediction using convolutional neural networks. BMC bioinformatics. 2021;22(1):561.
53. Corvelo A, Hallegger M, Smith CW, Eyras E. Genome-wide association between branch point properties and alternative splicing. PLoS computational biology. 2010;6(11):e1001016.
54. Hutchinson EC, Curran MD, Read EK, Gog JR, Digard P. Mutational analysis of cis-acting RNA signals in segment 7 of influenza A virus. J Virol. 2008;82(23):11869-79.
55. de Wit E, Spronken MI, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. Virus research. 2004;103(1-2):155-61.
56. Turnbull ML, Wise HM, Nicol MQ, Smith N, Dunfee RL, Beard PM, et al. Role of the B allele of influenza A virus segment 8 in setting mammalian host range and pathogenicity. Journal of virology. 2016;90(20):9263-84.
57. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, et al. ImageJ2: ImageJ for the next generation of scientific image data. BMC bioinformatics. 2017;18(1):529.
58. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, et al. Function of alternative splicing. Gene. 2005;344:1-20.

59. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet. 2002;3(4):285-98.
60. Zhang K, Shang G, Padavannil A, Wang J, Sakthivel R, Chen X, et al. Structural–functional interactions of NS1-BP protein with the splicing and mRNA export machineries for viral and host gene expression. Proceedings of the National Academy of Sciences. 2018;115(52):E12218-E27.
61. Yángüez E, Nieto A. So similar, yet so different: selective translation of capped and polyadenylated viral mRNAs in the influenza virus infected cell. Virus research. 2011;156(1-2):1-12.
62. Marión RM, Aragón T, Beloso A, Nieto A, Ortín J. The N-terminal half of the influenza virus NS1 protein is sufficient for nuclear retention of mRNA and enhancement of viral mRNA translation. Nucleic acids research. 1997;25(21):4271-7.
63. Hale BG, Randall RE, Ortín J, Jackson D. The multifunctional NS1 protein of influenza A viruses. Journal of general virology. 2008;89(10):2359-76.
64. Reed R. The organization of 3'splice-site sequences in mammalian introns. Genes & development. 1989;3(12b):2113-23.
65. Smith CW, Chu TT, Nadal-Ginard B. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Mol Cell Biol. 1993;13(8):4939-52.
66. Gog JR, Afonso Edos S, Dalton RM, Leclercq I, Tiley L, Elton D, et al. Codon conservation in the influenza A virus genome defines RNA packaging signals. Nucleic Acids Res. 2007;35(6):1897-907.
67. Coolidge CJ, Seely RJ, Patton JG. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. Nucleic acids research. 1997;25(4):888-96.
68. Poon LL, Pritlove DC, Fodor E, Brownlee GG. Direct evidence that the poly (A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. Journal of virology. 1999;73(4):3473-6.
69. Nao N, Yamagishi J, Miyamoto H, Igarashi M, Manzoor R, Ohnuma A, et al. Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutinin. MBio. 2017;8(1):10.1128/mbio. 02298-16.
70. Sharp CP, Thompson BH, Hoque AF, Diebold O, Tesla B, Kurian D, et al. Understanding off-target growth defects introduced to influenza A virus by synonymous recoding. RNA. 2025;31(11):1557-74.
71. Reed R, Maniatis T. Intron sequences involved in lariat formation during pre-mRNA splicing. Cell. 1985;41(1):95-105.
72. Ruskin B, Greene JM, Green MR. Cryptic branch point activation allows accurate in vitro splicing of human β-globin intron mutants. Cell. 1985;41(3):833-44.
73. Li Q, Jiang Z, Ren S, Guo H, Song Z, Chen S, et al. SRSF5-Mediated Alternative Splicing of M Gene is Essential for Influenza A Virus Replication: A Host-Directed Target Against Influenza Virus. Adv Sci (Weinh). 2022;9(34):e2203088.
74. Zhang M, Zharikova D, Mozdzanowska K, Otvos L, Gerhard W. Fine specificity and sequence of antibodies directed against the ectodomain of matrix protein 2 of influenza A virus. Molecular immunology. 2006;43(14):2195-206.
75. Hom N, Gentles L, Bloom JD, Lee KK. Deep mutational scan of the highly conserved influenza A virus M1 matrix protein reveals substantial intrinsic mutational tolerance. Journal of virology. 2019;93(13):10.1128/jvi. 00161-19.
76. Inglis S, Brown C. Differences in the control of virus mRNA splicing during permissive or abortive infection with influenza A (fowl plague) virus. Journal of general virology. 1984;65(1):153-64.

77. Bourmakina SV, García-Sastre A. The morphology and composition of influenza A virus particles are not affected by low levels of M1 and M2 proteins in infected cells. J Virol. 2005;79(12):7926-32.
78. Takeda M, Pekosz A, Shuck K, Pinto LH, Lamb RA. Influenza a virus M2 ion channel activity is essential for efficient replication in tissue culture. Journal of virology. 2002;76(3):1391-9.
79. Watanabe T, Watanabe S, Ito H, Kida H, Kawaoka Y. Influenza A virus can undergo multiple cycles of replication without M2 ion channel activity. Journal of virology. 2001;75(12):5656-62.
80. Park YW, Katze MG. Translational control by influenza virus: Identification of cis-acting sequences and trans-acting factors which may regulate selective viral mRNA translation. Journal of Biological Chemistry. 1995;270(47):28433-9.
81. Burgui I, Aragon T, Ortín J, Nieto A. PABP1 and eIF4GI associate with influenza virus NS1 protein in viral mRNA translation initiation complexes. Journal of General Virology. 2003;84(12):3263-74.
82. Wang W, Cui Z-Q, Han H, Zhang Z-P, Wei H-P, Zhou Y-F, et al. Imaging and characterizing influenza A virus mRNA transport in living cells. Nucleic acids research. 2008;36(15):4913-28.
83. Burset M, Seledtsov IA, Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic acids research. 2000;28(21):4364-75.
84. Abril JF, Castelo R, Guigó R. Comparison of splice sites in mammals and chicken. Genome research. 2005;15(1):111-9.
85. Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. Genome packaging in influenza A virus. Journal of general virology. 2010;91(2):313-28.
86. Dadonaite B, Gilbertson B, Knight ML, Trifkovic S, Rockman S, Laederach A, et al. The structure of the influenza A virus genome. Nature microbiology. 2019;4(11):1781-9.
87. Chua MA, Schmid S, Perez JT, Langlois RA. Influenza A virus utilizes suboptimal splicing to coordinate the timing of infection. Cell reports. 2013;3(1):23-9.