Vaccine-Induced Covid-19 Mimicry Syndrome

Kowarz E\(^1\)*, Krutzke L\(^2\)*, Külp M\(^1\), Streb P\(^1\), Larghero P\(^1\), Reis J\(^1\), Bracharz S\(^1\), Engler T\(^2\), Kochanek S\(^2\), Marschalek R\(^1\)

* contributed equally

\(^1\) Institute of Pharmaceutical Biology/DCAL, Goethe-University of Frankfurt, Biocenter, Max-von-Laue-Str. 9, 60438 Frankfurt/Main, Germany;

\(^2\) Department of Gene Therapy, Ulm University, Helmholtz Str. 8/1, 89081 Ulm, Germany

Conflict of interest: all authors declare no conflict of interest

Author’s contribution: EK, LK, SK and RM designed experiments, EK, LK, MK, PS, PL, JR, SB and ET performed experiments and evaluated the data; SK, LK and RM wrote and corrected the manuscript.

corresponding author:
Prof. Dr. Rolf Marschalek
Inst. Pharmaceutical Biology
University of Frankfurt
Max-von-Laue-Str. 9
60438 Frankfurt/Main, Germany
Tel: +49-69-798-29647; Fax: +49-69-798-29662
E-mail: Rolf.Marschalek@em.uni-frankfurt.de
Abstract (207)

To fight the Covid-19 pandemic caused by the RNA virus SARS-CoV-2 a global vaccination campaign is in progress to achieve the immunization of billions of people mainly with adenoviral vector- or mRNA-based vaccines, all of which encode the SARS-CoV-2 Spike protein. In some rare cases, cerebral venous sinus thromboses (CVST) have been reported as a severe side effect occurring 4 to 14 days after the first vaccination and were often accompanied by thrombocytopenia. Besides CVST, splanchic vein thromboses (SVT) and other thromboembolic events have been observed. These events only occurred following vaccination with adenoviral vector-based vaccines but not following vaccination with mRNA-based vaccines. Meanwhile, scientists have proposed an immune-based pathomechanism and the condition has been coined Vaccine-induced Immune Thrombotic Thrombocytopenia (VITT). Here, we describe an unexpected mechanism that could explain thromboembolic events occurring with DNA-based but not with RNA-based vaccines. We show that DNA-encoded mRNA coding for Spike protein can be spliced in a way that the transmembrane anchor of Spike is lost, so that nearly full-length Spike is secreted from cells. Secreted Spike variants could potentially initiate severe effects when binding to cells via the ACE2 receptor. Avoiding such splicing events should become part of a rational vaccine design to increase safety of prospective vaccines.

Introduction

The Covid-19 pandemic, starting in the last months of 2019 in Wuhan (China) and caused by the RNA virus SARS-CoV-2, so far (4 Jan 2022) has resulted in more than 293.0 Mio infections and more than 5.46 Mio deaths (data source: https://www.worldometers.info/coronavirus/). The world has faced the fastest development and production of vaccines ever, resulting in a total of four vaccines that have received conditional market authorizations by the regulatory bodies: the mRNA vaccines from BioNTech/Pfizer (BTN162b2/Comirnaty; EMA: 21.12.20) and Moderna (mRNA-1273/ Spikevax; EMA: 6.1.21), as well as the adenoviral vector-based vaccines from AstraZeneca (AZD1222/ChAdOx1-S/Vaxzevria; EMA: 29.1.21) and Janssen (Ad26.COV2.S; EMA: 11.3.21). All four vaccines encode slightly different forms of the SARS-CoV-2 Spike glycoprotein, which mediates virus binding to the host cell membrane and entry via ACE2 and TMPRSS2, respectively (1).

While vaccination with BioNTech and Moderna mostly causes only mild and typical immediate vaccination side effects, severe side effects (Thromboses, TTS) were first observed with Vaxzevria in Europe (in March 2021). The most severe side effects entailed rare events of thrombocytopenia combined with cerebral venous sinus thrombosis (CVST). The higher than expected occurrence of CVST led to a rapid suspension of the Vaxzevria vaccination campaign in several European countries (11-14th March 2021: Denmark, several northern European countries, Thailand, Ireland; 15th March 2021: Germany, Italy, France and Spain).
The latest data from the security surveillance of the Paul-Ehrlich-Institute in Germany (report dated 23.12.2021; https://www.pei.de/DE/newsroom/dossier/coronavirus/sicherheitsbericht-covid-19-impfstoffe-aktuell.html) revealed still a significant risk increase for the AstraZeneca vaccine compared to the other vaccines (see Table 1). Thrombocytopenia and Immunothrombocytopenia (ITP) have been observed in 314 cases after 96.6 Mio vaccination doses with Comirnaty (3.3 cases per mio injections), in 28 cases with Spikevax after 10.5 Mio vaccination doses with Comirnaty (2.6 cases per mio injections), in 269 with Vaxzevria after 12.7 Mio vaccination doses cases (21.2 cases per mio injections), and in 23 cases with Ad26.COV2.S after 3.4 Mio vaccination doses cases (6.6 cases per mio injections). Thrombosis-related death cases for these 4 vaccines were 74, 6, 81 and 12, respectively. This indicates that thromboembolic events associated with case mortality were in the range of 21.4% - 52% (see Table 1).

Thrombosis with thrombocytopenia syndrome (TTS) has been observed for these 4 vaccines in 36, 5, 200 and 24 patients, indicating a relative risk of 0.4, 0.5, 15.7 and 6.9 cases per million injections. These thromboembolic events in combination with thrombocytopenia share similarities with another condition, “Heparin-induced thrombocytopenia” (HIT). The term “Vaccine-induced Immune Thrombotic Thrombocytopenia” (VITT) was coined for this vaccine-induced condition, after finding out that autoantibodies against the platelet factor 4 (PF4) could be causally linked to it (2,3). However, despite intensive research activities by many laboratories the underlying initiating mechanisms for the observed side events were not yet fully understood.

At the molecular delivery mechanism of the approved vaccines, there are fundamental differences between mRNA- and adeno viral vector-based vaccines. The mRNA vaccines encode, as the naming implies, the spike sequence as a modified RNA sequence encapsulated in lipid nanoparticles. Upon intramuscular injection, the nanoparticles are taken up by the cells, where the mRNA cargo is released into the cytosol. Since both mRNA vaccines encode a signal peptide, the process of protein translation occurs directly at the rough endoplasmatic reticulum (ER) and results in membrane-anchored Spike proteins. Further post-translational modifications occur in the ER and Golgi apparatus (glycosylation) from where the Spike protein trimer is transported in vesicles to the outer cell membrane.

In contrast, adeno viral vector-based vaccines deliver the spike sequence as a codon-optimized DNA sequence instead of a RNA sequence. Vaxzevria is based on a genetically modified chimpanzee adenovirus type Y25, while the Janssen vaccine is based on a human adenovirus type 26 (HAdV-D26), both of which are replication incompetent. Upon transduction of the host cell and disassembly of the vector particle, the linear double stranded DNA is released and enters the nucleus. Subsequently, viral DNA-encoded genes become transcribed using the host cell transcription machinery (4). In the nucleus and before the mRNA is exported to the cytosol for translation (see above), the resulting RNA molecule is subject to the same post transcriptional RNA processing machinery as endogenous primary transcripts, including the possibility of splicing.
SARS-CoV-2 is a single-stranded RNA virus of the betacoronavirus family that naturally has a purely cytoplasmic replication cycle. Therefore, coronaviruses have not been exposed to any evolutionary pressure that would have an impact on biological processes taking place in the nucleus. Almost all nuclear encoded genes contain introns that become eliminated from precursor RNA transcripts by the splice machinery (5). Defined RNA consensus sequences are recognized by RNA/protein complexes (spliceosome), which consist of nearly 100 proteins and specific U-RNAs. Noteworthy, splice events are influenced by exonic splice enhancer sequences (ESE) and exonic splice silencer sequences (ESS), which bind splice-activating or -suppressing proteins. During this evolutionary conserved splice process, a so-called "branch A"-nucleotide is performing a nucleophilic attack at a G•G-dinucleotide of the 5'-splice consensus sequence G•GUNNGU. The resulting 3'-OH of the final exonic G nucleotide then makes a second nucleophilic attack at the G•G-dinucleotide of the 3'-splice consensus sequence YYYYYCAG•G. To this end, 2 exons become fused, and the intronic sequence is being eliminated. For each gene, this process guarantees that all introns flanked by splice donor and acceptor sites are eliminated in order to produce the final mRNA that encodes a specific protein coding sequence.

When using cloned cDNA in molecular biology or biotechnology, no intronic sequences are usually present, because cDNA synthesis is made from already spliced mRNA molecules. However, in case of the SARS-CoV-2 virus, the cloned cDNA for the Spike open reading frame was not coming from a nuclear encoded gene, rather from an RNA virus that replicates solely in the cytosol of infected cells. Therefore, such a cDNA exhibits potential splice donor and acceptor sites just by chance (~130 cryptic splice signal sequences in the total SARS-CoV-2 virus sequence), because no negative selection has ever taken place during the evolution of Corona virus to eliminate such sequences. Therefore, we hypothesized that nuclear splicing of DNA-encoded Spike RNA sequences, might result in spliced, and thus, truncated spike protein variants with unknown fate and function. As mentioned above, this type of splicing would be quite artificial, because an RNA virus does per se not contain any intronic sequences in its various open reading frames. Thus, all such splicing events are derived from so-called cryptic splice sites. This is unrelated to the mechanism of alternative splicing, which could be functionally attributed to certain genes of our genome. Cryptic splicing occurs only if strong splice sites are present, and thus, these cryptic splice sites will be used equally in different cell types or tissues.

To address this question of cryptic splicing, we analyzed nuclearly transcribed wildtype and codon-optimized Spike genes in a specially designed splice reporter vector system. We also investigated splicing of Spike RNAs following transduction of representative cells with a Spike-expressing human adenovirus type 5 vector and the two approved adenoviral vector-based vaccines ChAdOx1-S and Ad26.COV2.S. Here, we show that splicing of DNA-encoded SARS-CoV-2 Spike RNA molecules takes place at positions as they were predicted by in silico analysis. These splice events resulted in Spike protein variants that lack the transmembrane anchor, thus may become secreted due to the nature of the secretory pathway. Codon-
optimization performed to enhance expression levels further augmented the abundance of splicing and was most prominently observed for the ChAdOx1-S vector. In addition, we analyzed the production of Spike protein and its secretion to the outside medium. Implications of our findings for the safety of vector-based vaccines and future consequences for the design of vector-based vaccines for RNA viruses will be discussed.

**Results**

**In silico analyses reveal potential splice sites in adenoviral vector-encoded spike sequences**

The Spike open reading frame of the Wuhan SARS-CoV-2 isolate was tested for potential splice sites by using the available SpliceRover online tool (6). In addition, we used the online Alternative Splice Site Predictor (ASSP) tool to confirm the predicted splice sites (data not shown). Both servers identified several dozens of potential splice sites in the 3,822 nucleotide long open reading frame of the Spike gene. Therefore, we decided to filter these hits by their individual scores to identify those, which were used with a higher likelihood (wildtype Spike gene: 6 splice donor (SD) sites, 5 splice acceptor (SA) sites, see Fig. 1 and Table 2A).

Similarly, we performed the same in silico analysis also for codon-optimized Spike open reading frames in three different adenovirus vector systems. First, these were an experimental and codon-optimized variant of the original Wuhan sequence, cloned into a human adenovirus type 5 vector backbone (Ad5.S). Second, the Spike sequence of ChAdOx1-S was derived by sequencing of adenoviral DNA isolated from Vaxzevria, while, third, the Spike sequence of Ad26.COV2.S was obtained directly from Janssen (kindly provided by Roland Zahn). Analyses revealed that codon optimization resulted in an increased number of potential splice sites in all three adenoviral-encoded spike sequences (Ad5.S: 12 Splice donor (SD) sites, 10 splice acceptor (SA) sites; ChAdOx1-S: 13 SD and 5 SA sites, Ad26.COV2.S: 9 SD and 10 SA sites, see Tables 2B-D) compared to the open reading frame of the wildtype Wuhan SARS-CoV-2 Spike sequence.

**Significant splicing of codon-optimized spike sequences**

First, we analyzed potential splice events within the open reading frame of wildtype Spike, which contains, according to our prediction, 6 potential splice donor sites. Since these splice signal sequences does not occur like in normal genes where every splice donor site is located upstream of a splice acceptor site, both flanking an intronic sequenc, such strong splice donor site are “unsaturated” and usually give rise to cryptic splice reactions in cis or in trans to potential splice acceptors (7). To this end, we generated three splice traps in which the spike sequence was inserted upstream of a perfect splice acceptor site, which again was fused in three different reading frames (0, +1, +2) to a full-length Luciferase gene without start codon. In this system Luciferase activity is only detected if splicing occurs. Respective constructs
were cloned into a sleeping beauty transposon vector (pSBbi::Spike-Luc-0/+1/+2-GP) and used for stable transfection of HEK293T cells.

Analyses of all three splice trap constructs in Luciferase assay revealed that the “+2 construct” exhibited the strongest intracellular Luciferase activity (1,291 relative light units), while the “0 construct” resulted in a much lower activity (341 relative light units). The activity determined for the +1 construct was near the values for non-transfected HEK293T cells, which served as negative controls (see Fig. 2, panel I.). We also analyzed the medium of the cultured cells and observed that small amounts of Luciferase activity were present also extracellularly. This amount increased overtime and mimicked relative ratios of the three constructs observed intracellularly (see Fig. 2, panels II., III.). Due to the design of the three splice traps, Luciferase activity can only be detected when splicing occurs. Thus, these initial Luciferase experiments confirmed that splicing of the spike sequences was taking place. If splicing occurs between a SD in the Spike open reading frame and the SA in the splice trap, the C-terminally located transmembrane domain of the spike protein serving as the membrane anchor, will get lost. Thus, resulting truncated spike proteins may become secreted.

We also cloned the three codon-optimized Spike genes of the adenoviral vectors (Ad5.S, ChAdOx1-S, Ad26.COV2.S) into the three versions of the Luciferase splice trap (0/+1/+2) and analyzed the production of Spike-Luciferase fusion proteins. Here, we only analyzed intracellular Luciferase activity (see Fig. 2, panels IV. - VI.). All three codon-optimized Spike genes derived from the different adenoviral vectors displayed a strong production of Spike-Luciferase fusion proteins that was also enhanced compared to wildtype Spike sequence constructs. Thus, all codon-optimized Spike genes displayed a higher splicing activity than the wildtype Spike gene sequence. The highest activity of 15,840 relative light units was detected with the reading frame +1 of the ChAdOx1-S construct. The lowest level of Luciferase activity was found in the Ad26.COV2.S construct.

We also performed Western Blot experiments to determine steady-state levels of intracellularly produced adenovirus vector-encoded Spike protein variants und Spike-Luciferase fusion proteins. Cell lysates and supernatants of cell lines were used for Spike-immunoprecipitation experiments using protein G beads and anti-FLAG antibodies for capturing. Subsequently, samples were analyzed by Western Blot analysis using anti-Spike and anti-Luciferase antibodies. As shown in Fig. 3, for all analyzed spike splice traps strong signals of proteins with sizes of 95-250 kDa were detected in cell lysates with both antibodies, confirming spike-luciferase fusion proteins of different lengths. Cell lysate and medium supernatant of each investigated cell line displayed several protein bands migrating at molecular weights as predicted by the major splice events in each of the investigated cell lines (Ad5.S 195,1 and 195,3 kDa; ChAdOx1-S 195,1 and 195,3 kDa; Ad26.COV2.S 196,4 kDa; all without glycosylation).
Verification of various splice products derived from wildtype and codon-optimized Spike open reading frames

To validate these findings at the molecular level, we analyzed RNA from cell lines transfected with the three wildtype Spike-encoding splice trap constructs. Potential splice events were visualized by RT-PCR experiments using a set of four Spike primers pointing 3'-prime against one Luciferase primer pointing 5'-prime. All deviating PCR bands were analyzed by Sanger sequencing (see Fig. 4, agarose gels). As shown in Fig. 4, we observed several splice reactions from predicted splice sites (e.g. SD1-3), but also internal splice reactions that lead to frame-shift events. We also found two splice events downstream of the ACE2 binding domain resulting again in a fusion between Spike and Luciferase.

To exclude that observed occurring splicing of the Spike mRNA was an artifact due to the use of artificial splice trap constructs with a very potent splice acceptor site at the 5'-end of the Luciferase reading frame, we repeated the analysis using the actual adenoviral vectors Ad5.S, ChAdOx1-S and Ad26.COV2.S. Here the Spike sequence is embedded in the adenoviral DNA genome, directly upstream of the adenoviral pIX. Being part of the viral capsid, pIX has mainly structural functions as a “cement” protein increasing stability of the viral particle, but also has additional functions in the viral life cycle (8). The 5'-non-coding sequence of the pIX transcription unit of the three different adenovirus types contains a highly conserved splice acceptor site behind the pIX promoter and upstream of the pIX ATG start codon (Ad5.S pIX at nucleotide -14; ChAdOx1-S pIX at nucleotide -4; Ad26.COV2.S pIX at nucleotide -5). In wildtype adenoviruses this site is naturally used as splice acceptor during processing of the adenoviral E1B RNA. To investigate potential splicing between the Spike sequence and the adenoviral pIX HeLa and HepG2 cells were transduced with the adenoviral vectors Ad5.S, ChAdOx1-S or Ad26.COV2.S. Preceding uptake assays for each cell line had been performed to determine required multiplicities of infection (MOI) of the different adenoviral vectors to achieve comparable cellular transduction efficiencies with the different adenoviral vector types (Suppl. Fig. S2). Two days post transduction, total RNA was extracted from the cells and analyzed by RT-PCR using universal pIX reverse primer to detect splice events between various locations of the Spike-coding RNA and the pIX transcript. As shown in Figures 5-7, all three Spike genes of the different adenoviral vector backbones displayed splice events to the adenoviral pIX transcript. Noteworthy, although all Spike genes contain a poly(A) signal sequence that should initiate the polyadenylation process, in all cases we observed transcriptional read-through allowing the detected splice events to occur. Only recurrently identified splice events are displayed in Figures 5-7. The Ad5.S Spike gene displayed the predominant splice events at positions 506, 1821, 3610*, 3614 and 3641* to the SA within the pIX transcript of the vector backbone (numbers with an asterisk were not predicted by SpliceRover). The ChAdOx1-S Spike gene displayed the predominant splice events at positions 506, 1049*, 1914*, 3610*, 3614 and -298* and -84*. The latter two splice events involved a SD located in the bovine growth hormone (BGH) poly(A) signal 298 and 84 nucleotides upstream of the pIX start codon. Finally, the Ad26.COV2.S Spike gene displayed weak splice
events occurring at nucleotide 541*, 959*, 1184, 1221*, 3641* and -83*. Also here one splice event involved a SD in the simian virus 40 (SV40) poly(A) signal sequence located -83 nucleotides upstream of the pIX start codon.

**Predominant and cell type-dependent prevalence of ChAdOx1-S spike RNA splicing**

To validate the previous analyses of observed splicing between the adenoviral vector-encoded spike and adjacent adenoviral protein pIX transcripts, we performed additional PCR on cDNA samples of HeLa cells transduced with equal efficiencies (Figures 8, 9) with Ad5.S, ChAdOx1-S and Ad26.COV2.S vectors. Using forward primers that bind to the spike sequence and reverse primers that bind to the pIX sequences revealed for all three vectors read-through events with DNA fragment lengths of 1802 bp, 2037 bp and 1865 bp for Ad5.S, ChAdOx1-S and Ad26.COV2.S, respectively (Figure 9, green arrows). While in Ad26.COV2.S-transduced HeLa cells no truncated cDNA fragments were detected, PCR amplification of Ad5.S- and ChAdOx1-S-transduced HeLa cells samples showed additional DNA bands of shorter lengths, confirming splicing at different positions within the read-through sequences. Interestingly, sequencing of the shortest DNA fragment of Ad5.S-transduced samples (Figure 9, red arrow) revealed that splicing occurred in frame from the predicted SD site upstream of the transmembrane (TM) domain of the spike protein to the SA site upstream of the pIX start codon (Fig. 5, splice site at nt 3610), resulting in a potentially secreted fusion protein. Similar results were obtained from sequencing of the shortest DNA fragment of ChAdOx1-S-transduced samples (Figure 9, A: blue arrow): splicing took place between the predicted SD site upstream of the TM of the spike protein (Fig. 6, splice site at nt 3610) to the SA site upstream of the pIX start codon (Figure 9, F). However, this splicing event resulted in the generation of a stop codon, thus no fusion- but though secreted, truncated proteins would be translated due to the lacking TM domain. Sequencing of the most abundant DNA fragment in ChAdOx1-S samples (Figure 9, white arrow) revealed splicing from the predicted SD site within the BGH poly(A) tail to the SA site upstream of the pIX start codon (Fig. 6, splice site at nt -298).

Comparing the total amount of PCR-generated DNA indicates significant higher spike-pIX read-through copy numbers (and subsequent potential splicing) for Ad5.S-transduced HeLa samples than ChAdOx1-S-transduced HeLa samples (note even 2.5-fold less volume of PCR samples was loaded for Ad5.S; Figure 9, A). Since expression cassettes of both vectors are under the control of the identical hCMV promoter, this data suggests that the BGH Poly(A) of ChAdOx1-S is more potent than the SV40 Poly(A) of Ad5.S to prevent read-through transcription into the pIX gene.

Interestingly, the use of cells of other tissue origin in particular the human hepatocyte cell line HepG2 and the murine C2C12 myocytes, demonstrated that the abundance of splicing events has been cell type-dependent. In cDNA samples of Ad5.S-transduced C2C12 cells only the full-length read-through fragment was detected –
indicating no splicing at all (Figure 9, C: green arrow). In contrast, for ChAdOx1-S-transduced C2C12 cells almost exclusively the BGH-pIX splice product was detected (Figure 9, C: white arrow and Figure 9, E). Moreover, in hepatocytes the ratio between full-length and spliced products was again different for Ad5.S- or ChAdOx1-S-transduced cells (Figure 9, B). While PCR patterns of Ad5.S-transduced HepG2 cells looked the same as with HeLa cells, in ChAdOx1-S-transduced HepG2 cells the Spike-pIX splice product was the predominantly detected fragment (Figure 9, B: blue arrow and Figure 9, F). Since these splice products lack the TM-coding sequence, the resulting protein will likely be secreted.

**Significant amounts of splice events detected in genome-wide RNA-Seq data of ChAdOx1-transduced cells**

To broaden our perspective regarding the occurrence of splicing upon cell transduction by ChAdOx1 from a pIX-focused to a genome-wide level, we also analyzed publicly available RNA-sequencing data sets of ChAdOx1-S-transduced MRC-5 cells (9). These data sets were based on nanopore sequencing data, which by nature are rather error-prone and thus analysis is complicated. Therefore, we first filtered the data sets for reads containing the ChAdOx1-S Spike gene. In a second step, we both looked for the presence of pIX sequences and blasted all Spike sequence reads against the human genome. The result of such an exemplary data mining experiment is shown in Fig. 10, where the analysis of the data set derived from MRC-5 cells at day 72 after transduction with ChAdOx1-S is displayed. Out of millions of reads, 9,632 reads could be attributed to the codon-optimized Spike gene. As schematically depicted in Fig. 10, the Spike open reading frame is flanked at the 5’-end by the CMV promoter with a small intron and at the 3’-end by the BGH poly(A) signal and the pIX gene, respectively. A 93 nucleic acid sequence coding for the signal sequence of the tissue plasminogen activator (tPA) protein replaces the original start codon of the Spike gene. In this analysis, we identified, within the 9,632 reads containing Spike nucleotide sequences, a total of 120 trans-spliced mRNAs, containing either host cell gene exons (fusing a total of 1 until 18 exons) fused to either the genuine splice acceptor of the CMV promoter (n=8) or the internal Spike splice acceptor sites (n=47). Similarly, we found Spike sequences fused to exons of host cell genes (between 1-28 exons). These trans-splicing events occurred either from the Spike open reading frame (n=32, 32 different splice donor sites) or from the identified BGH splice donor sites (n=33, 1 splice donor site). Exemplarily, we have identified the following genes: RSP, MSPG2, GPN2, CCN1, MRPL9, SORT1, UROS, TMEM35B, PRDX, ADAR, IFITM1, MMP1, DUD, PKM, ALDOA, MT2A, DD42, etc. A total of 50 Spike-pIX fusions were found as well (8 splice donor sites), while 79 reads were found from the BGH splice donor to the splice acceptor upstream of the pIX coding sequence as it has been observed with HeLa, HepG2 and C2C12 cells. Although the frequency of these splicing event was rather low compared to normal Spike reads (1.24-1.33% versus 97.41%) their identification was unexpected. For comparison, we also determined the number of reads for 2 housekeeping genes, ABL1 and GAPDH, respectively: ABL1 reads were only 44, while GAPDH reads were...
4,586. Thus, the amount of 249 reads that displayed aberrant forms of Spike should be considered as significant.

**Discussion**

Here, we provide experimental evidence that adenovirus-based vaccines, designed to express the SARS-CoV-2 Spike protein from a cloned, codon-optimized cDNA may result in unwanted splicing events, if the presence of functional splicing sites in the coding sequence is not avoided during vector design.

SARS-CoV-2 is an RNA virus with an infectious cycle that takes place entirely in the cytoplasm. Thus, there has never been an evolutionary selection against presence of splice donor and acceptor sites in its gene sequences. While mRNA-based vaccines that are functionally only active in the cytoplasm, transcripts from adenoviral vectors are generated in the cell nucleus, where also post-transcriptional RNA processing steps - including splicing – is taking place. Thus, the findings shown and discussed here have implications for DNA-encoded vaccine development not only against SARS-COV-2 but also against other RNA viruses.

In this study, we experimentally validated our hypothesis that Spike protein variants lacking the natural membrane anchor - thereby being subjected to secretion – are generated from adenoviral vaccine vectors as a consequence of nuclear RNA splicing events. We showed, (1) that in a Spike/splice reporter system use of *in silico* predicted splice donor sites within the wildtype Spike gene resulted in Spike-Luciferase protein fusion events, (2) that in this system splice reactions lead to secreted protein variants due to the splicing-dependent loss of the transmembrane membrane domain, (3) that two of the three evaluated adenoviral vectors that encode a spike gene that is codon-modified for improved expression exhibited more and stronger splice donor sites than the viral wildtype sequence, and (4) that following transcriptional read-through, splicing occurred to a strong splice acceptor in the adenoviral pIX transcript, going also along with loss of the transmembrane anchor of Spike.

We wish to point out that in our experimental studies we have investigated a distinct spectrum of splice events, since we used primers binding to either Luciferase or the adenoviral pIX coding sequence for PCR amplification. Alternative splice events, e.g. internal splice events (in-frame, out-of-frame) were mostly cloned and sequenced by chance. The analysis of provided RNA-Seq data after cells were infected with the AstraZeneca vaccine gave insight into additional trans-splicing events with host cell genes. Viral RNA genes evolutionary contain splice site consensus sequences due to a missing counter selection. Thus, we conclude that nuclear expression of such RNA virus-derived genes in mammalian cells may result in unwanted splicing events with host cell genes that should be considered when developing new vaccines in the future.

Our results clearly indicate that Spike splice reactions occur and that secreted soluble Spike protein variants are generated. Western blot analyses from total cells and supernatant thereof revealed a series of Spike protein variants that strongly vary
in their molecular weight, indicative for the previously analyzed splicing events. Soluble Spike protein has been described to cause adverse effects, e.g. a strong inflammatory response on endothelial cells (10-13). Obviously, severe life-threatening thromboembolic events occur due to infection of recipient cells by wildtype SARS-CoV-2. Even pseudoviruses with Spike protein located at their surface may cause strong inflammatory reactions in tissues and endothelial cells, indicating a potential risk of soluble Spike protein to cause severe side effects when systemically present in the vascular system (14).

According to our results, soluble Spike variants could potentially explain the occurrence of rare but severe events observed in particular after vaccination with Vaxzevria. Noteworthy, the Ad26.COV2.S vaccine carries fewer splice donor sequences, especially SD_{506} and SD_{3614} are absent (see Table 2B-D), which are the strongest predicted splice donor sites in the Ad5.S and ChAdOx1-S encoded Spike sequences (see Figure 1). This was further confirmed by the observation that in contrast to Ad5.S and ChAdOx1-S, no significant amounts of spliced RNA molecules upon cell transduction with Ad26.COV2.S were generated, This may explain the ~3-fold lower incidence for thromboembolic side effects with the Janssen vaccine when compared to the Vaxzevria vaccine (see Table 1). It may also explain the ~2-fold lower risk for TTS with the Janssen vaccine.

Based on our data we pose the hypothesis that soluble Spike via its receptor-binding domain (RBD) will bind to ACE2 and anti-Spike antibodies generated during immunization will recognize Spike variants, bound to the endothelial cell surface. Inflammatory reactions could locally occur through different mechanisms including antibody dependent cell-mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC), both of which could serve as starting point for thrombus formation. In ADCC, mainly NK cells are recruited to surface-bound antibodies via Fc-gamma receptors (CD16 or CD32) and cause cell lysis after their activation. In CDC, anti-Spike antibody coated endothelial cells may recruit and activate complement that may result in platelet activation via binding of C3 and terminal complement complex on the surface of platelets and thrombin/serotonin secretion (15). By this means, thromboses may occur in any site of the human body where endothelial cells express ACE2.

Why should such events occur in the cerebral dural venous sinus system? It is conceivable that due to the absence of venous valves bi-directional and posture-dependent blood flow in the dural venous sinus the residence time of soluble Spike protein is increased and the possibility to bind to endothelial cells expressing ACE2 enhanced. This could increase the probability of adverse immune reactions at places of slow blood flow.

Recently the vaccine-induced immune thrombotic thrombocytopenia (VITT) syndrome has been described, suggesting a complex inflammatory process. Related to Heparin induced thrombocytopenia (HIT) it has been reported that the induction of anti-PF4 autoantibodies and platelet activation in a subset of patients, may result in the observed combination of thrombocytopenia and thromboses mainly in large vessels including cerebral venous sinus and (Figure 11, left side). Additionally, it is
possible that protein impurities from the vaccine production process may be involved in the generation of a pro-inflammatory status (16). Here we propose a third pathomechanism as it is depicted in Figure 1, right side. Based on our splicing data, membrane-anchored and secreted soluble Spike protein variants are produced after vaccination. Secreted Spike protein variants are disseminated throughout the vascular system and may concentrate at endothelial cells expressing the ACE2 at their surface. These ACE2-bound Spike will then become a target of anti-Spike antibodies, generated after vaccination with the potential to result in ADCC- and/or CDC-mediated inflammatory reactions. It may also cause the already described NET reaction of neutrophils that has been causally linked to thromboembolic events (17-19). Also the rarely observed capillary leak syndrome (Clarkson-disease; 20) could potentially be explained by such a mechanism.

The distinguishing features of the three different adenovirus vectors Ad5.S, ChAdOx1-S and Ad26.COV2.S have been summarized in Table 3. Based on the analysis of the design of the vaccines and their splice behavior of the encoded transcripts as described in this report we have identified several aspects that should be considered for future adenovirus vaccine design and development.

We have two different data sets. One data set has been created by large scale sequencing of splice events of all 3 adenoviral vector systems after transduction in Hela and HepG2 cell lines (Figures 5-7), where the different splice events of all three vectors have been investigated by simply analyzing a large number of PCR fragments through direct sequencing or subcloning and sequencing. The second data set is based on experiments, where all 3 adenoviral vectors were analyzed for their splice activity after adjusting the MOI’s in the Hela, HepG2 and C2C12 cell lines (Figure 8). In the latter, read-through transcripts have been investigated for the activity of the different poly(A) signal sequences (BGH vs SV40), but also for the amount of splice events in that particular transcript (Figure 9). Transcriptional read-through obviously is a precondition for the generation of splice events between the Spike-coding sequence and the pIX gene encoded by the adenoviral vector genome. Very little transcriptional read-through was observed with ChAdOx1-S that can be linked to the BGH poly(A) signal used in this construct. Apparently, the SV40 poly(A) signal is not as efficient in termination of the transcript as we observed with Ad5.S. In these cases, differences were observed for splice events which may point to the fact that tissue-specific splicing may account as well as for some differences in the creation of soluble Spike protein variants.

Based on our findings, we suggest that the Spike open reading frames – wildtype or codon-optimized for expression - in DNA-based vaccines should always be analyzed for the presence of strong splice donor (SD) sites. If possible, strong SD should be avoided to prevent unintended splice. We note that the Janssen Ad26.COV2.S vaccine has been modified to reduce the number of strong SDs, while the Ad5.S and ChAdOx1-S vectors still contain strong SDs, which is reflected in a large number of detectable splicing events, while only very few splicing events were detected in the Ad26.COV2.S vaccine.
Very little is known about fate and tissue-distribution of adenovirus-based vaccine in humans. In mice, it has been observed that following intramuscular injection significant amounts of adenovirus vectors end up in hepatocytes, where expression of the transgene occurs (21). It is likely that the Chimpanzee ChAdOx1-S is also a hepatotropic adenovirus, since it is based on the Chimpanzee adenovirus Y25 (22), which belonged to a group of isolates from chimpanzees associated with outbreaks of viral hepatitis (23). If transduction of hepatocytes would occur in (some of the) vaccines after intramuscular injection in a significant amount, it is conceivable that larger amounts of Spike proteins, having lost the transmembrane domain by splicing, will be secreted into the systemic circulation via the hepatic veins.

Taken together, from our work it appears that for DNA-encoded sequences (especially when derived from RNA viruses), most importantly, strong SDs should be avoided and strong poly(A) signals should be used for RNA termination in expression constructs. This is especially important when adenoviral vectors are used as transport vehicles. Usually, transgenes sequences are embedded in the adenoviral genome in the gap of the deleted E1-region, located directly upstream of the adenoviral pIX with a strong SA site. Alternatively, while most transgenes in adenoviral vectors have a “left-to-right” orientation, also an inversion of the expression cassette would avoid splicing events to occur between the transgenic coding sequence and the viral pIX transcript. Taking this into account, chances of splicing will be strongly reduced or null.

We strongly emphasize the consideration of these aspects for the rational design of safer future adenoviral vaccine vectors and also for adenoviral vector-based gene transfer approaches in general.

Material and Methods

Splice site prediction

The SARS-CoV-2 Spike nucleic acid sequences of of the Wuhan SARS-CoV-2 strain (NC_045512.2) and codon-optimized Spike sequences encoded by the three investigated adenoviral vectors Ad5.S, ChAdOx1-S, and Ad26.COV2.S were analyzed for the presence of splice donor and acceptor sites using the state-of-the-art SpliceRover online tool (bioit2.irc.ugent.be/rover/splicerover). Sites with a score below 0.15 were excluded from further analysis (6). The Alternative Splice Site Predictor (ASSP) online tool was used as second algorithm for potential splice site identification (24).

Cloning of the open reading frame of SARS-CoV-2 Spike protein

cDNA prepared from the RNA of the Wuhan SARS-CoV-2 strain (NC_045512.2) was kindly provided by Tuna Toptan and Marek Widera (Institute of Medical Virology, University Hospital Frankfurt am Main, Goethe-University, Frankfurt am Main, Germany). The cDNA was used for PCR experiments and for cloning of the open
reading frame of the Spike gene following amplification using the following primers:

CoV-2_S.Flag.Sfi.F 5'-aGGCCTCTGAGGCCaccatggattacaaggatgacgacgataagatgttt-
gttttctgttttattgaccatgtctct-3',

5'-aGGCCTGACAGGCCttatgtgtaatgtaatttgactcctttgagcac-3'. Resulting amplimers were
cloned into the Topo II vector system and appropriate clones were validated by
sequence analyses. The final clone was digested with SfiI and cloned into the
Sleeping Beauty transposon vector system pSBTet-GP (GFP/Puromycin; 25). The
recombinant Spike was N-terminally modified with a FLAG-tag.

Development of a Spike/splice reporter system

The cloned wildtype Spike open reading frame was used to establish an additional
splice reporter system. The Spike/splice reporter system consists of the full-length
Spike gene including the stop codon, followed by an intronic sequence that contains
the necessary branch-A nucleotide, as well as a perfect splice acceptor site fused in
different reading frames (0, +1, +2) to a full-length Luciferase gene (without
start codon). In this system Luciferase activity is only detected if splicing occurs.
Thus, 3 independent Spike/splice reporter constructs (pSBbi::Spike-Luc-0/+1/+2-GP)
were, cloned into the integrating Sleeping Beauty transposon vector pSBbi-GP (25).
Additionally, isolated adenoviral DNAs of Ad5.S, ChAdOx1-S, and Ad26.COV2.S
were used as templates to PCR-amplify the 3822 bp codon-optimized open reading
frames of the Spike sequences. These PCR fragments were also cloned into the 3
Luciferase reporter gene constructs pSBbi::LUC-0/+1/+2-GP. The resulting Sleeping
Beauty transposon vectors were sequence-validated and subsequently stably
integrated into the genome of HEK293T cells for RT-PCR and Luciferase reporter
experiments. The ChAdOx1-S spike sequence constructs was cloned without the tPA
leader sequence (+31 aa) in order to make all 3 different constructs (Ad5.S,
ChAdOX1-S, Ad26.COV2.S) equal in length. The AGA codon at position 94-96 of the
original AZ vector was thus converted into an ATG start codon. This way, all tested
construct had an open reading frame of 3822 nucleotides.

Stable transfection experiments

All splice reporter constructs were stably transfected into HEK293T cells and
selected for 3-5 days with 2 µg/ml Puromycin. Cells were used for Luciferase
experiments or for the isolation of total RNA (Qiagen) in order to investigate potential
splice events by RT-PCR.

Adenoviral vectors

Ad5.S particles produced in N52.E6 cells (25), had a physical titer of 3.35 x 10^8 VP/µl
and were dissolved in 50 mM HEPES, 150 mM NaCl, 10% glycerol, pH 7.4.
According to the manufacturer, the ChAdOx1-S vaccine from AstraZeneca (lot
number ABV9317) had a physical titer of 1 x 10^8 VP/µl and particles were dissolved
in 10 mM histidine, 7.5% sucrose (w/v), 35 mM NaCl, 1 mM MgCl2, 0.1% polysorbate
80 (w/v), 0.1 mM EDTA and 0.5% EtOH (w/v). According to the manufacturer
Ad26.COV2.S vaccine from Janssen (lot number XD955 and 21C10-01) had a titer of $1 \times 10^8$ VP/µl and particles were dissolved in citric acid monohydrate (0.14 mg), trisodium citrate dihydrate (2.02 mg), ethanol (2.04 mg), 2-hydroxypropyl-β-cyclodextrin (HBCD) (25.50 mg), polysorbate-80 (0.16 mg), sodium chloride (2.19 mg).

**Production and purification of Ad5.S vectors**

Ad5.S vector particles used in this study are E1- and E3-deleted replication-incompetent vector particles based on human adenovirus species C type 5 (based on GenBank AY339865.1, Δ441–3522 and Δ28123-30813). The CMV-promoter controlled SARS-CoV2 Spike expression cassette with a SV40 poly(A) signal is based on the GenBank sequence YP_009724390.1, codon-optimized for Homo sapiens and was inserted in the E1 region. Particles were produced in E1-complementing N52.E6 cells (26) and purified by CsCl gradient ultracentrifugation as described earlier (27).

In brief: $4 \times 10^8$ cells were transduced with MOI 300 from stock solution. 48 h post-transduction cells were harvested, resuspended in 3 ml 150 mM NaCl, 50 mM HEPES, pH 7.4 and lysed by freeze/thaw cycles. Cell debris was separated by centrifugation at 2,000 x g for 10 min, and supernatants were layered on a CsCl step gradient (density bottom: 1.41 g/ml; density top: 1.27 g/ml, 50 mM HEPES, 150 mM NaCl, pH 7.4) and centrifuged at 176,000 x g for 2 h at 4 °C. Vector particles were aspirated and further purified by a consecutive continuous CsCl gradient (density: 1.34 g/ml, 50 mM HEPES, 150 mM NaCl, pH 7.4) and centrifuged at 176,000 x g for 20 h at 4 °C. Vector particles were aspirated and desalted by size exclusion chromatography using PD10 columns (GE Healthcare). Physical vector titers were determined by optical density measurement at 260 nm (28).

**Isolation of adenoviral DNA from vector stock solution**

Adenoviral DNA was isolated from 200 µl vector stock solutions of Ad5.S, ChAdOx1-S, and Ad26.COV2.S using the QIAamp DNA Mini Kit (250). Eluted DNA was precipitated with EtOH and pellets resuspended in 10 µl 10 mM Tris buffer, pH 8.0.

**Cell lines**

HEK293T cells were maintained in DMEM Low Glucose medium mixed with 10% (v/v) FBS, and supplemented with 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. HeLa (human cervical cancer), HepG2 (human hepatocyte carcinoma) and C2C12 (murine muscle myoblast) were cultivated in MEM, alphaMEM or DMEM medium respectively, all of which were supplemented with 10% FBS and 1% antibiotics. Cells were grown at 37°C in 5% CO$_2$ and a relative humidity of 95%. Cells were passaged twice a week except for C2C12, which were passaged three times a week. For HeLa and HepG2 cells adenoviral vector transduction was done 24 h after cell seeding, whereas C2C12 cells were incubated for 72 h after seeding to let them grow dense and start differentiation into myocytes. Cell lines were obtained from ATCC (HEK293T) or DMSZ (Hela, HepG2, C2C12) and are
tested for their cell identity by STR profiling as provided by company. In addition, these cell stocks were tested for mycoplasma infections (as not already done by the company), before being expanded in cell culture for 14 days and subsequently frozen as cell stocks in liquid nitrogen. For all cell culture experiment, such cell stocks were thawed and used in cell culture for a maximum of several weeks.

Cellular uptake assays with different adenoviral vectors

1 x 10^5 HeLa, HepG2 and C2C12 cells were transduced with indicated physical multiplicity of infection (MOI) of Ad5.S, ChAdOx1-S or Ad26.COV2.S. Cells were incubated with the vectors for 2 h at 37 °C before cells were harvested and total DNA was isolated using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) according to the manufacturer instructions. DNA concentrations were determined by optical density measurement at 260 nm. Intracellular adenoviral genome copy numbers in 20 ng total DNA were analyzed by quantitative real time PCR. Primers used for Ad5.S and ChAdOx1-S samples: fw.: 5'-TAGACGATCCCTACTGTACG-3'; rv.: 5'-GGAAATATGACTACGTCCGG-3'. Primers used for Ad5.S and Ad26.COV2.S samples: fw.: 5'-CAGGACGCTCGGAGTACCTGAG-3'; rv.: 5' GGGGCCACCGGGGTTT-3'. DNA was dissolved in 2 µl and mixed with 10 µl SYBR Green (Kapa Biosystems), 0.4 µl 10 pmol/µl of each forward and reverse primer in a total volume of 20 µl. Thermocycles: first cycle: 10 min at 95 °C; 40 cycles: 30 sec 95 °C, 30 sec 60 °C, 8 sec 72 °C; followed by a final cycle of 10 min at 72 °C. Assays were performed in biological and technical triplicates.

Analysis of SARS-CoV-2 spike RNA splicing after adenovirus vector-mediated cell transduction

Adenoviral cell transduction assays were performed using HeLa, HepG2, and C2C12 cells. To achieve equivalent cell transduction efficiencies with adenovirus vectors of different strains, HeLa cells were transduced with MOI 1000 for Ad5.S, MOI 3000 ChAdOx1-S and MOI 6000 for Ad26.COV2.S. HepG2 cells were transduced with MOI 1000 for Ad5.S, MOI 3000 for ChAdOx1-S, and MOI 10,000 for Ad26.COV2.S. C2C12 cells were transduced with MOI 1000 for both Ad5.S and ChAdOx1-S (see Figure 8). Cells were incubated with the vectors for 48 h at 37 °C, before RNA was isolated using RNeasy Plus Mini Kit (Qiagen) followed by cDNA generation using Maxima H Minus cDNA Synthesis (ThermoScientific) according to the manufacturers’ instructions. cDNA was used for RT-PCR (see section RT-PCR) or 2 µl cDNA was used for Taq polymerase PCR in a total volume of 25 µl. Forward primer binding to the spike sequence: 5'-CAAGGACTTCGCGGGCTTCAA-3’, used for all three vectors. Reverse primer sequences binding to the adenoviral pIX sequence: 5'-CCCATCACATTCTGACCGCA-3’ for Ad5.S and ChAdOx1-S; 5'-TGCTGTCGAGCGGAGTTTC-3’ for Ad26.COV2.S. Thermocycles: 3 min at 95 °C; 40 cycles: 30 sec 95 °C, 30 sec 55 °C, 2 min 68 °C. 8 µl of Ad5.S and 20 µl of ChAdOx1-S and Ad26.COV2.S PCR products were separated on agarose gels, DNA bands were excised, purified with gel extraction kit (Qiagen), cloned using the pCR.
2.1-TOPO TA cloning kit (ThermoScientific) and sequenced using primers 5’-CAGGAAACAGCTATGAC-3’ and 5’-GTAAAACGAGGCCAG-3’. Assays were performed in biological triplicates.

**RT-PCR experiments**

For RT-PCR experiments a series of primers were used. To test for splice events in the 3 reporter cell lines, we used the following primer sequences: S.0233.F 5’-GGTTTGATAACCCTGTCCTACCA-3’, S.1027.F 5’-AACGCCACCAAGATTTGCATC-3’, S.1712.F 5’-AACAAGATGGCTGCTCGT-3’, and the reverse primer Luc4008.R 5’-GTCCACCTCAGATGTG-3’. RT-PCR experiments were carried out under stringent conditions, and PCR fragments deviating from expected PCR bands were analyzed by Sanger sequencing.

For the analysis of RNA derived from adenoviral vector transduced HeLa and HepG2 cells we developed a series of universal primers. These were the following primers: 506univ.F 5’-GCGAGTTCCAGTTCTGCAACG-3’, 1810univ.F 5’-CAGACACTGGAAATCCTGGGACTAC-3’, 3610.F 5’-GCAGGATCTAGCTGCT-3’ and the reverse primer plXuniv.R 5’-CGTCAGAATGTGATGGAGACG-3’. RT-PCR experiments were carried out under stringent conditions, and PCR fragments deviating from expected PCR bands were analyzed by Sanger sequencing. Thermocycles: first cycle: 2 min at 94 °C; 35 cycles: 20 sec 95 °C, 30 sec 60 °C, 40 sec 72 °C; followed by a final cycle of 3 min at 72 °C.

**Luciferase assays**

Luciferase assays were performed using HEK293T cells that were stably transfected with the Sleeping Beauty Luciferase reporter gene constructs that exhibits a constitutive GFP expression cassette. 5,000 cells per well were seeded in black glass bottom 96-well plates (Greiner BioOne). After two days, GFP fluorescence (RFU) was measured with the Varioskan Flash Plate-reader (Thermo Fischer Scientific). Subsequently, D-Luciferin (final concentration of 187.5 µg/ml) was added with the automatic dispense system, and Luciferase activity was determined (RLU). A normalization procedure was used to express the RLU/RFU ratio to account for intercellular differences. All measurements were carried out with n=6, while luciferase activity in medium was measured once at two different time points (t=0 and t=48h).

For this purpose, 200 µl medium was centrifuged for 5 minutes at 800 x g and 100 µl supernatant was used to determine Luciferase activity.

**Immunoprecipitation and Western Blot Analyses**

Cells were lysed with immunoprecipitation buffer (150 mM NaCl, 10 mM Tris, 1 mM EDTA, 1 mM EGTA, 1% Triton-X-100, 0.5% NP-40, protease inhibitor cocktail (Roche Diagnostics)) for 30 min under constant agitation at 4 °C. Cell lysates were separated from cell debris by centrifugation for 5 min at 2000 rpm (376 x g) and 4°C.
Lysates were quantified using Pierce BCA Protein Assay Kit (Thermo Scientific) according to the manufacturer’s protocol and adjusted to 40 µg/µl.

The amount of 200 µl of cell lysate or cell culture supernatant (3 days) was pre-cleared with 25 µl protein G magnetic beads (New England Biolabs) in low-binding tubes (Biozym Scientific) for 1h under constant agitation at 4 °C. After magnetic separation the supernatant was incubated with 25 µl anti-FLAG® M2 magnetic beads (Sigma-Aldrich #M8823) for 2 h under constant agitation at 4 °C. Beads were then washed three times with 0.5 ml IP buffer and resuspended in 20 µl 2x Lämmli buffer. For elution the precipitate beads were incubated for 5 min at 70 °C and beads magnetically separated.

10 µl of supernatant was electrophoretically separated using Mini-PROTEAN TGX stain-free gradient gels 4-15% (Biorad) and color protein standard broad range P7719S (NEB). Proteins were blotted onto low fluorescence western PVDF membrane (Abcam) in Towbin buffer (25 mM Tris, 192 mM Glycin, 15% methanol, 0.01% SDS, pH=8.3) at 110 V for 70 min. Membranes were blocked using 5% BSA in TBS-T for 1 h at room temperature. Primary antibody incubation was performed over night at 4 °C. Membranes were washed with 5% BSA in TBS-T four times prior to incubation with HRP-conjugated secondary antibody for 1 h at room temperature. After four wash steps with 5% BSA in TBS-T chemiluminescence detection was performed using Clarity™ Western ECL Substrate (Biorad) and ChemiDoc™ XRS+ system (Biorad).

The following antibodies were used: α-Firefly Luciferase pAb (Abcam #ab21176) diluted 1/1,000, α-SARS-CoV-2 Spike pAb (Sino Biologicals #40589-T62) diluted 1/1,000, α-rabbit IgG-HRP (Abcam #ab6721) diluted 1/10,000.

Acknowledgement

We thank all lab members for fruitful discussions and their help. We also want to thank Prof. Owen Williams for critical reading of the manuscript. We also thank Dr. Tuna Toptan and Dr. Marek Widera for providing us with the cDNA of SARS-CoV-2 (Wuhan isolate). We are also grateful to Dr. Roland Zahn (Janssen) for providing the Spike gene sequence of the Janssen vaccine. This work was supported by a funding grant of the Corona Task Force of the Goethe-University to RM.

Competing interests: The authors declare no conflict of interests


Table and Figure legends

Table 1: Safety data from the German Paul-Ehrlich Institute (PEI), dated from 23.12.2021
All data for all 4 vaccines were directly retrieved from the security report. Data for number of all vaccinations, total cases with complications, anaphylaxie cases, myo-/pericarditis cases, cases with thromboses associated with thrombozytopenia (TTS), cases with thromboses combined with immunethrombocytpenia (ITP), and cases with Guillain-Barré-Syndrom (GBS) were listed. In addition, for each category the case numbers per 1 million injections are given (e.g. the absolute risk for the development of myo/pericarditis after a Biontech vaccination is 12.9 cases per million injections). All red numbers differ significantly from the numbers of the other vaccines.

Table 2: Splice donor site prediction
The open reading frame of the SARS-CoV-2 Spike protein was analyzed by the Splice site Predictor online tool SpliceRover (http://bioit2.irc.ugent.be/rover/splicerover). This online tool allows prediction of splice donor and splice acceptor sites. Predicted splice donor sites are provided in Table 1 for the SARS-CoV-2 wildtype Spike gene (A), for the codon-optimized Spike gene derived from the experimental Ad5.S vector, and the FDA/EMA-approved vector-based vaccines from AstraZeneca (Vaxzevria, ChAdOx1-S) and Janssen/J&J (Ad26.COV2.S), respectively. The amino acid coordinates of Spike protein domains S1, S2 and the minimal ACE2 binding domain are indicated.

Table 3: Comparison of all 3 adenoviral vector-based constructs
Based on our analysis of three different adenoviral vectors, there were differences between the vaccines. The displayed table analyzes the use of promoter, the different poly(A) elements and changes inside the sequences of Spike e.g. additional sequences, presence or deletion of the Furin cleavage sequence, the addition of 2 prolines to stabilize of the prefusion configuration of Spike and site-directed mutations that should avoid splicing.

Figure 1: Splice site prediction within the Spike open reading frames
Splice site prediction was carried out by using SpliceRover. Splice donor sites are given in red, splice acceptor sites in green. SpliceRover calculates splice sites with a score between 0 to 1, but only splice sites with >0.15 were displayed. Sites were displayed in 4 colors as indicated. Splice sites were numbered with “0”, “+1” or “+2”, to indicate how the open reading frame is disrupted. Therefore, all splice events between splice sites with identical numbers will be in-frame, while all splice reactions between unequal numbers will result in out-of-frame fusions. Below, the protein domain structure of Spike is displayed to explain, which domains are being deleted by splice events (NTD: N-terminal domain, RBD: (ACE2) receptor binding domain, RRAR: Furin cleavage site, HR1: Heptad repeat1, CH: central helix, CD: connector domain, TM: transmembrane domain). White asterisks mark three splice donor sites that are present in the codon optimized Spike reading frame of the Vaxzevria but not of the Janssen vaccine.

Figure 2: Splice trap Luciferase experiments in HEK293T cells
The three different splice traps were cloned into the pSBbi-GP vector. This vector encodes two polycistronic transcripts, one encoding GFP and Puromycin resistance (from left to right), while the other encodes the full-length Spike gene connected to an artificial intron and an ATG-deleted Luciferase gene. Splice events to the Luciferase transcript will result in “in-frame” and “out-of-frame” Spike-Luciferase fusion transcripts. HEK293T cells were stably transfected with the constructs and resulting Luciferase activity was measured. RLU: relative light units; RFU: relative fluorescence units. Middle panels I-III: The three different splice trap constructs encode the wildtype Spike gene. Luciferase expression was measured intracellularly and extracellularly 3 day post transfection. Lower panels IV-VI: The three different splice trap constructs encode the codon-optimized spike sequences derived from either the adenoviral vector Ad5.S, ChAdOx1-S or Ad26.COV2.S. Luciferase expression was measured intracellularly.

Figure 3: Splice trap Western blot experiments in HEK293T cells
Spike and Spike-fusion proteins were concentrated with the help of immunoprecipitation beads from all cell lysates and supernatants of HEK293T cells stably transfected with the splice trap constructs encoding the codon-optimized Spike sequences derived from the adenoviral vectors Ad5.S, ChAdOx1-S or Ad26.COV2.S. Protein bound to these beads were used to perform Western blot experiments, which are displayed. Bead eluates were separated by SDS-PAGE, blotted and proteins were detected with anti-Spike or anti-Luciferase antibodies.

Figure 4: Splice events after transfection of HEK293T cells
The wildtype SARS-CoV-2 Spike gene - cloned into the three splice trap constructs - was stably transfected in HEK293T cells, RNA was isolated and investigated by RT-PCR experiments for splicing events. Appropriate PCR bands were cut-out from the gel and either directly sequenced or cloned for sequencing. The major splice events discovered are displayed.

Figure 5: Splice events after Ad5.S transduction of Hela cells
HeLa or HepG2 cells were transduced with the adenoviral vectors Ad5.S and RNA was isolated 48 h post transduction. Three universal primers binding at three different locations within the Spike genes and one universal reverse primer binding to the adenoviral pIX that is located directly adjacent to the Spike gene sequence were used to investigate splicing events. The resulting DNA fragments are shown in the agarose gel on the left. Appropriate PCR bands were excised from the gel and either directly sequenced or cloned for sequencing. The predominately detected splice events are displayed on the right.

Figure 6: Splice events after ChAdOx1-S transduction of Hela or HepG2 cells
HeLa or HepG2 cells were transduced with the adenoviral vector Ad5.S and RNA was isolated 48 h post transduction. Three universal primers binding at three different locations within the Spike genes and one universal reverse primer binding to the adenoviral pIX that is located directly adjacent to the Spike gene sequence were used to investigate splicing events. The resulting DNA fragments are shown in the agarose gel on the left. Appropriate PCR bands were excised from the gel and either directly sequenced or cloned for sequencing. The predominately detected splice events are displayed on the right.
Figure 7: Splice events after Ad26.COV2.S transduction of HeLa or HepG2 cells

HeLa or HepG2 cells were transduced with the adenoviral vectors Ad5.S and RNA was isolated 48 h post transduction. Three universal primers binding at three different locations within the Spike genes and one universal reverse primer binding to the adenoviral pIX that is located directly adjacent to the Spike gene sequence were used to investigate splicing events. The resulting DNA fragments are shown in the agarose gel on the left. Appropriate PCR bands were excised from the gel and either directly sequenced or cloned for sequencing. The predominately detected splice events are displayed on the right.

Figure 8: Analysis of required MOI of different vector strains to achieve equal transduction efficiencies

HeLa, HepG2 and C2C12 cells were transduced with indicated physical multiplicity of infection (MOI) of Ad5.S, ChAdOx1-S or Ad26.COV2.S. Cells were incubated with the vectors for 2 h at 37°C and harvested. Total DNA was isolated and intracellular adenoviral genome copy numbers were analyzed by quantitative real time PCR.

Figure 9: Cell type-dependent and predominant splicing of the ChAdOx1.S spike sequence

(A) HeLa, (B) HepG2 and (C) C2C12 cells were transduced with equal efficiencies with the adenoviral vectors Ad5.S, ChAdOx1-S and Ad26.COV2.S encoding the codon-optimized spike sequences. Total RNA was extracted 48 h post transduction. PCRs using forward primers that bind to the spike sequence and reverse primers that bind to the adjacent downstream pIX in the adenoviral DNA genome were used for PCR amplification of generated cDNAs. Separation of PCR amplicons by agarose gel electrophoresis. Green arrow: full length read-through fragments; Red arrow: Spliced fragment of the Ad5.S-encoded spike sequence resulting in a potentially secreted spike fusion protein. White arrow: Spliced fragment of the ChAdOx1-S-encoded spike sequence resulting in a Spike-pIX fusion protein. Blue arrow: Spliced fragment of the ChAdOx1-S-encoded spike sequence resulting in a potentially secreted spike protein. (D) Sequencing result of DNA fragment marked with the red arrow. (E) Sequencing result of DNA fragment marked with the white arrow. (F) Sequencing result of DNA fragment marked with the blue arrow.

Figure 10: Trans-splicing events of Vaxzevria in MRC-5 cells

RNA Seq data were kindly provided by David A. Matthews, University of Bristol. RNA Seq data derived from MRC-5 cells after transduction with the Vaxzevria vaccine. RNA was isolated after 72 h and sequenced in depths. We used this data set to investigate potential splice events visible in this cell line. Total numbers and locations of detected trans-splice and pIX-splice events within the Spike sequence and the adjacent pIX sequence are graphically depicted. As displayed, out of 9.632 Spike reads 120 reads were obtained from trans-splicing events from or to host mRNAs. 129 reads were found that demonstrated splicing upstream of the pIX gene. We also determined the amount of 2 housekeeping genes (ABL1 and GAPDH) to analyze their read numbers. In comparison to the housekeeping genes vector-based transcription was quite strong and superseded the amount of GAPDH transcripts.

Figure 11: Predicted pathological disease mechanism
I. predicts VITT mechanism that is accompanied by the thrombocytopenic situation. II. pro-inflammatory events which might be caused by vaccine impurities. III. proposed mechanism that soluble Spike protein variants are generated by distinct splice events. Soluble Spike protein variants which are able to bind to ACE2 expressing endothelial cells may trigger immunological mechanisms that result in thromboembolic events (CVST and SVT; estimated 1 in 80,000 vaccinated persons in Europe). RBD: receptor binding domain; TM: transmembrane domain; SD: splice donor; SA: splice acceptor, pA: poly-A sequence; ??? fusion to unknown sequences.
Table 1: Safety data from the German Paul-Ehrlich Institute (PEI), dated from 23.12.2021

<table>
<thead>
<tr>
<th>Company</th>
<th>Biontech/Pfizer</th>
<th>Moderna</th>
<th>Oxford/AZ</th>
<th>Janssen/J&amp;J</th>
</tr>
</thead>
<tbody>
<tr>
<td>#vaccinations</td>
<td>96,606,131</td>
<td>10,576-131</td>
<td>12,703,030</td>
<td>3,462,557</td>
</tr>
<tr>
<td>#complications</td>
<td>113,792</td>
<td>28,289</td>
<td>46,325</td>
<td>7,758</td>
</tr>
<tr>
<td>cases per Mio inj.</td>
<td>1,178</td>
<td>2,675</td>
<td>3,647</td>
<td>2,241</td>
</tr>
<tr>
<td>Anaphylaxie</td>
<td>550</td>
<td>55</td>
<td>101</td>
<td>10</td>
</tr>
<tr>
<td>cases per Mio inj.</td>
<td>5,7</td>
<td>5,2</td>
<td>8,0</td>
<td>2,9</td>
</tr>
<tr>
<td>Myo/Perikarditis</td>
<td>1,245</td>
<td>309</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cases per Mio inj.</td>
<td>12,9</td>
<td>29,2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TTS</td>
<td>36</td>
<td>5</td>
<td>200</td>
<td>24</td>
</tr>
<tr>
<td>cases per Mio inj.</td>
<td>0,4</td>
<td>0,5</td>
<td>15,7</td>
<td>6,9</td>
</tr>
<tr>
<td>ITP</td>
<td>314</td>
<td>28</td>
<td>269</td>
<td>23</td>
</tr>
<tr>
<td>cases per Mio inj.</td>
<td>3,3</td>
<td>2,6</td>
<td>21,2</td>
<td>6,6</td>
</tr>
<tr>
<td>GBS</td>
<td>140</td>
<td>14</td>
<td>112</td>
<td>48</td>
</tr>
<tr>
<td>cases per Mio inj.</td>
<td>1,4</td>
<td>1,3</td>
<td>8,8</td>
<td>13,9</td>
</tr>
<tr>
<td>death cases</td>
<td>295</td>
<td>20</td>
<td>201</td>
<td>21</td>
</tr>
<tr>
<td>cases per Mio inj.</td>
<td>3,1</td>
<td>1,9</td>
<td>15,8</td>
<td>6,1</td>
</tr>
</tbody>
</table>
### Table 2: Splice donor site prediction

#### A: splice donor site prediction in wildtype Spike ORF

<table>
<thead>
<tr>
<th>Position</th>
<th>pot. splice donor sites</th>
<th>Score</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>454-473</td>
<td>TGAGATGGGA•GTGAGTTCAG</td>
<td>0,187</td>
<td>+1</td>
</tr>
<tr>
<td>541-560</td>
<td>GAAACACGGA•GGATTTCGAA</td>
<td>0,151</td>
<td>+1</td>
</tr>
<tr>
<td>894-911</td>
<td>GAAACAAGT•GTACGTGGAA</td>
<td>0,565</td>
<td>+1</td>
</tr>
<tr>
<td>1323-1342</td>
<td>TGATTTCTAG•GTGAGAAGTA</td>
<td>0,357</td>
<td>0</td>
</tr>
<tr>
<td>1906-1925</td>
<td>TATTTACAG•GTTCCTATTFT</td>
<td>0,274</td>
<td>+1</td>
</tr>
<tr>
<td>1996-2015</td>
<td>ATTTGTCGCA•GTATATGCGC</td>
<td>0,707</td>
<td>+1</td>
</tr>
<tr>
<td>3296-3317</td>
<td>GCACACACTG•GGTTTGAAGA</td>
<td>0,160</td>
<td>+2</td>
</tr>
</tbody>
</table>

#### B: splice donor site prediction in codon-optimized Spike ORF of Ad5.S

<table>
<thead>
<tr>
<th>Position</th>
<th>pot. splice donor sites</th>
<th>Score</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>497-516</td>
<td>GCACCTTCGA•GTACGTGCC</td>
<td>0,987</td>
<td>+2</td>
</tr>
<tr>
<td>1175-1194</td>
<td>TCACAAAGTG•GTACCGCGAG</td>
<td>0,187</td>
<td>+2</td>
</tr>
<tr>
<td>1209-1228</td>
<td>GGAGATGCAA•GTGCAGCGAG</td>
<td>0,162</td>
<td>0</td>
</tr>
<tr>
<td>1812-1831</td>
<td>CTCCAACAG•GTGGCGTGTC</td>
<td>0,540</td>
<td>0</td>
</tr>
<tr>
<td>2331-2350</td>
<td>CACCAAGGAG•GTTCGGCCGC</td>
<td>0,204</td>
<td>0</td>
</tr>
<tr>
<td>2949-2968</td>
<td>ACTGGAGCGA•GTGGAGCGCG</td>
<td>0,318</td>
<td>0</td>
</tr>
<tr>
<td>2961-2980</td>
<td>GGAAGCGGAG•GTGCAGCGCG</td>
<td>0,151</td>
<td>0</td>
</tr>
<tr>
<td>3083-3102</td>
<td>AGATGTCTGA•GTGTTTATG</td>
<td>0,318</td>
<td>+2</td>
</tr>
<tr>
<td>3296-3315</td>
<td>GCACCCATTG•GTGTCGACC</td>
<td>0,388</td>
<td>+2</td>
</tr>
<tr>
<td>3452-3471</td>
<td>AACTGGATAA•GTACCTTATG</td>
<td>0,443</td>
<td>+2</td>
</tr>
<tr>
<td>3555-3574</td>
<td>GCTGAACAGA•GTGGCAAGA</td>
<td>0,217</td>
<td>0</td>
</tr>
<tr>
<td>3605-3624</td>
<td>AACTGGGGGA•GTACAGCGAG</td>
<td>0,980</td>
<td>+2</td>
</tr>
</tbody>
</table>

#### C: splice donor site prediction in codon-optimized Spike ORF of ChAdOx1-S

<table>
<thead>
<tr>
<th>Position</th>
<th>pot. splice donor sites</th>
<th>Score</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>497-516</td>
<td>GCACCTTCGA•GTACGTGCC</td>
<td>0,986</td>
<td>+2</td>
</tr>
<tr>
<td>1012-1030</td>
<td>TCCGGCGAG•GTGTCGATG</td>
<td>0,266</td>
<td>0</td>
</tr>
<tr>
<td>1175-1194</td>
<td>TCACAAAGTG•GTACCGCGAG</td>
<td>0,177</td>
<td>+2</td>
</tr>
<tr>
<td>1209-1228</td>
<td>GGAGATGCAA•GTGCAGCGAG</td>
<td>0,188</td>
<td>0</td>
</tr>
<tr>
<td>1812-1831</td>
<td>CTCCAACAG•GTGGCGTGTC</td>
<td>0,541</td>
<td>0</td>
</tr>
<tr>
<td>2331-2350</td>
<td>CACCAAGGAG•GTTCGGCCGC</td>
<td>0,215</td>
<td>0</td>
</tr>
<tr>
<td>2949-2968</td>
<td>ACTGGAGCGA•GTGGAGCGCG</td>
<td>0,398</td>
<td>0</td>
</tr>
<tr>
<td>2961-2980</td>
<td>GGAAGCGGAG•GTGCAGCGCG</td>
<td>0,273</td>
<td>0</td>
</tr>
<tr>
<td>3083-3102</td>
<td>AGATGTCTGA•GTGTTTATG</td>
<td>0,503</td>
<td>+2</td>
</tr>
<tr>
<td>3296-3315</td>
<td>GCACCCATTG•GTGTCGACC</td>
<td>0,381</td>
<td>+2</td>
</tr>
<tr>
<td>3452-3471</td>
<td>AACTGGATAA•GTACCTTATG</td>
<td>0,245</td>
<td>+2</td>
</tr>
<tr>
<td>3555-3574</td>
<td>GCTGAACAGA•GTGGCAAGA</td>
<td>0,310</td>
<td>0</td>
</tr>
<tr>
<td>3605-3624</td>
<td>AACTGGGGGA•GTACAGCGAG</td>
<td>0,686</td>
<td>+2</td>
</tr>
</tbody>
</table>

#### D: splice donor site prediction in codon-optimized Spike ORF of Ad26.COV2.S

<table>
<thead>
<tr>
<th>Position</th>
<th>pot. splice donor sites</th>
<th>Score</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>563-582</td>
<td>ACCTGGCGCA•GTTCGTGTTC</td>
<td>0,150</td>
<td>+2</td>
</tr>
<tr>
<td>1175-1194</td>
<td>TCACAAAGTG•GTACCGCGAG</td>
<td>0,180</td>
<td>+2</td>
</tr>
<tr>
<td>1209-1228</td>
<td>GGAGATGCAA•GTGCAGCGAG</td>
<td>0,207</td>
<td>0</td>
</tr>
<tr>
<td>1812-1831</td>
<td>CACCAAGTACG•GTGGACGAG</td>
<td>0,547</td>
<td>0</td>
</tr>
<tr>
<td>2331-2350</td>
<td>CACCAAGGAG•GTTCGGCCGC</td>
<td>0,202</td>
<td>0</td>
</tr>
<tr>
<td>2961-2980</td>
<td>TGGAGCGGAG•GTGCAGCGCG</td>
<td>0,194</td>
<td>0</td>
</tr>
<tr>
<td>3083-3102</td>
<td>AGATGTCTGA•GTGTTTATG</td>
<td>0,495</td>
<td>+2</td>
</tr>
<tr>
<td>3296-3315</td>
<td>GCACCCATTG•GTGTCGACC</td>
<td>0,392</td>
<td>+2</td>
</tr>
<tr>
<td>3452-3471</td>
<td>AACTGGGCAA•GTACCTTAAAG</td>
<td>0,436</td>
<td>+2</td>
</tr>
</tbody>
</table>

---

**S1 ectodomain: encoded by nucleotides 1-2049 (aa 1-683)**

**S2 domain: encoded by nucleotides 2050 - 3822 (aa 684-1273)**

**ACE2 binding domain: encoded by nucleotides 998-19720 (aa 319-551)**
Table 3: Comparison of all 3 adenoviral vector-based constructs

<table>
<thead>
<tr>
<th>name</th>
<th>vector</th>
<th>promoter</th>
<th>additional sequences</th>
<th>terminator</th>
<th>RRAR</th>
<th>PP</th>
<th>splicing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad5.S</td>
<td>HAd5</td>
<td>CMV with Intron</td>
<td>-</td>
<td>SV40</td>
<td>present</td>
<td>no</td>
<td>+</td>
</tr>
<tr>
<td>ChAdOx1-S</td>
<td>ChAd</td>
<td>CMV with Intron</td>
<td>+tPA (31aa)</td>
<td>BGH</td>
<td>present</td>
<td>no</td>
<td>+</td>
</tr>
<tr>
<td>Ad26.COV2.S</td>
<td>HAd26</td>
<td>CMV w/o Intron</td>
<td>-</td>
<td>SV40</td>
<td>deleted</td>
<td>present</td>
<td>splice sites partially deleted</td>
</tr>
</tbody>
</table>

comment: upstream ATG

pos 682-685
pos 986/987
AZ >> J
Figure 3

<table>
<thead>
<tr>
<th>KDa</th>
<th>IP Lysat</th>
<th>supernatant</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ad5.S**
predicted by cloned splice events: +1 and +2

**ChAdOx1-S**
predicted by cloned splice events: +1 and +2

**Ad26.COV2.S**
predicted by cloned splice events: +2
Figure 4

Splice trap vector identified Luciferase splice events

Read through
0/1/2

Luciferase
SARS-CoV-2 Spike

Splice acceptor

S.0233.F xLuc.4008.R
S.1027.F xLuc.4008.R
S.1712.F xLuc.4008.R
S.2482.F xLuc.4008.R
ChAdOx1.S construct identified pIX splice events

Figure 6
Trans-splice activities with host genome

MRC-5 cells 72h

9,632 Total Spike Gene Reads

120 trans-splice events from or in genes of host = 1.24%
129 pIX splice events = 1.33%
9,383 normal SPIKE RNA = 97.43%

ChAdOx1

ABL1
44 reads

GAPDH
4,586 reads
vector based vaccine

adenoviral protein/DNA

auto-antibodies

Thrombocytopenia VITT

Vaccine impurities pro-inflammatory situation

Figure 11