



## Communications

# Homologous SV40 RNA *trans*-splicing: Special case or prime example of viral RNA *trans*-splicing?

Sushmita Poddar <sup>a</sup>, Joachim Eul <sup>b</sup>, Volker Patzel <sup>a,\*</sup>

<sup>a</sup> Department of Microbiology, Yong Loo Lin School of Medicine, National University of Singapore, Block MD4, Level 5, 5 Science Drive 2, Singapore 117597, Singapore

<sup>b</sup> INEIDFO GmbH, Weserstrasse 23, 12045 Berlin, Germany

## ARTICLE INFO

Available online 8 July 2014

## Keywords:

Alternative splicing  
RNA *trans*-splicing  
Viral RNA *trans*-splicing  
SV40  
HIV-1  
Adenovirus

## ABSTRACT

To date the Simian Virus 40 (SV40) is the only proven example of a virus that recruits the mechanism of RNA *trans*-splicing to diversify its sequences and gene products. Thereby, two identical viral transcripts are efficiently joined by homologous *trans*-splicing triggering the formation of a highly transforming 100 kDa super T antigen. Sequences of other viruses including HIV-1 and the human adenovirus type 5 were reported to be involved in heterologous *trans*-splicing towards cellular or viral sequences but the meaning of these events remains unclear. We computationally and experimentally investigated molecular features associated with viral RNA *trans*-splicing and identified a common pattern: Viral RNA *trans*-splicing occurs between strong cryptic or regular viral splice sites and strong regular or cryptic splice sites of the *trans*-splice partner sequences. The majority of these splice sites are supported by exonic splice enhancers. Splice sites that could compete with the *trans*-splicing sites for *cis*-splice reactions are weaker or inexistent. Finally, all but one of the *trans*-splice reactions seem to be facilitated by one or more complementary binding domains of 11 to 16 nucleotides in length which, however occur with a statistical probability close to one for the given length of the involved sequences. The chimeric RNAs generated via heterologous viral RNA *trans*-splicing either did not lead to fusion proteins or led to proteins of unknown function. Our data suggest that distinct viral RNAs are highly susceptible to *trans*-splicing and that heterologous viral *trans*-splicing, unlike homologous SV40 *trans*-splicing, represents a chance event.

© 2014 Poddar et al.

Viruses represent the smallest replicating specimens. Mammalian viruses are mainly composed of the viral DNA or RNA genome, protein coats protecting the genetic material, and in some cases lipid envelopes surrounding the coat in the extracellular stage. The size of a viral particle correlates with the size of its genome and for most viruses genome size is a common theme. In order to express a maximum number of proteins, viruses use an array of mechanisms which do not significantly enlarge the genomic size such as including the use of multiple promoters, alternative open reading frames, translational frame shifting, stop codon read-through, and alternative splicing. In the same sense, viruses use virus-encoded small interfering (si)RNAs and antisense transcription to regulate gene expression. Recently, we reported strong evidence that viruses, i.e. the simian virus 40 (SV40), do also recruit the mechanism of RNA *trans*-splicing to generate the full repertoire of protein functions [1]. In mammalian cells, RNA *trans*-splicing represents a special form of alternative splicing in which two distinct RNA transcripts are joined *in trans* with the help of the spliceosome leading

to the formation of chimeric RNA molecules which can code for novel functions or non-sense proteins. In the case of homologous SV40 RNA *trans*-splicing, two identical early viral transcripts are joined triggering the formation of a 100 kDa super tumour antigen (sT-ag). This 100 kDa protein was found to have superior cell transformation activity as compared with the regular 94 kDa large T antigen (T-ag). The SV40 *trans*-splice reaction was described to be supported by various SV40-intrinsic molecular helper functions which render this reaction highly efficient with about 50 to 70% of the viral pre-mRNA transcripts being involved in *trans*-splicing. Sequences of other viruses including HIV-1 and the human adenovirus type 5 were reported to be involved in *trans*-splicing as well [2,3]; however, these events have not been investigated in depth and the question as to whether these reactions represent chance events or are of relevance to the viruses remains open. We used SV40 RNA *trans*-splicing as a model and computationally screened sequences reported to be involved in heterologous viral RNA *trans*-splicing for common *trans*-splicing-associated molecular features and signatures.

As opposed to conventional *cis*-splicing, the efficiency of *trans*-splicing depends on the probability that two separate RNA precursor molecules encounter each other in the nuclear compartment. This probability increases with high sub-cellular RNA concentrations, e.g. in consequence of productive viral RNA transcription and/or when the

\* Corresponding author. Tel.: +65 6516 3318; fax: +65 6777 5720.

E-mail addresses: [sushmita@nus.edu.sg](mailto:sushmita@nus.edu.sg) (S. Poddar), [joachim.eul@snaflu.de](mailto:joachim.eul@snaflu.de) (J. Eul), [micvp@nus.edu.sg](mailto:micvp@nus.edu.sg) (V. Patzel).

**Table 1**  
Analyses of splice sites and binding domains.

Virus			Trans-splicing partner			Potential binding domains	Trans-splice product
Name	Splice information		Name	Splice information			
	Trans ss	Alt cis ss		Trans ss	Alt cis ss		
SV40 exon 2	5'cry U1 snRNA 6/9 bp, 14hb ESE 18 nts Sc 8.5/Sc 0.74	Not relevant	SV40 exon 2	3'reg 12/18 bp 4 sites, 6/6 bp, 13hb 18 nts 6.4/0.98	5'reg (wt) 5/9 bp, 12hb 10 nts 8.7/0.73  5'reg (mut) <sup>a</sup> 4/9 bp, 10hb 10 nts 8.7/0.73	10mer (22hb) 13mer (31hb)	Highly transforming 100 kD <sup>a</sup> super T antigen
Adeno-virus type 5 exon 2 of major late transcript (MLT)	3'reg U2AF65 16/18 bp U2 snRNA 5 sites, 4/6 bp, 9hb Sc 13.3/Sc 1.00	5'reg U1 snRNA 8/9 bp, 19hb ESE 8 nts 10/0.99	v-abl Abelson murine leukaemia viral proto-oncogene homolog 2 transcript variant a (exon 1A)	5'cry 6/9 bp, 15hb 11.8/0.95	3'reg 12/18 bp 3 sites, 3/6 bp, 7hb 8 nts 8.5/0.74	12mer (29hb)	No fusion protein
			Zinc Finger protein (ZNF) 41 exon 3	5'reg 9/9 bp, 22hb 9 nts 12.8/0.98	3'reg 17/18 bp 4 sites, 6/6 bp, 12hb 9 nts	11mer (27hb) 12mer (33hb)	Fusion protein, function unknown
			Armadillo protein P0071 5 UTR	5'cry 6/9 bp, 14hb	Not relevant	12mer (32hb) 12mer (30hb) 13mer (34hb) 13mer (33hb)	No fusion protein
			GTP-binding protein (RAB5) 1st coding exon	5'reg 6/9 bp, 14hb 7 nts	Sequence information not found	11mer (27hb) 11mer (27hb) 11mer (30hb) 11mer (29hb) 12mer (27hb)	Fusion protein, function unknown
			GAPDH exons 3, 5, 7	Exon 3: 5'reg 7/9 bp, 17hb 12.4/0.99	3'reg 14/18 bp 2 sites, 5/6 bp, 11hb 7.1/0.98	12mer (33hb) 12mer (30hb) 12mer (32hb) 12mer (31hb) 13mer (35hb) 13mer (35hb) 13mer (33hb) 15mer (40hb) 16mer (44hb)	Fusion protein, function unknown
				Exon 5: 5'reg 8/9 bp, 18hb 8 nts 12.2/0.96	3'reg 12/18 bp 2 sites, 5/6 bp, 12hb 5.9/0.98		
				Exon 7: 5'reg 4/9 bp, 10hb 7 nts 8.6/0.69	3'reg 12/18 bp 3 sites, 3/6 bp, 6hb 3.4/0.91		
			β-Actin exon 3	5'reg 9/9 bp, 22hb 8 nts 16.3/0.99	5'reg 12/18 bp 2 sites, 3/6 bp, 7hb 9.7/0.85	12mer (34hb) 13mer (38hb) 13mer (33hb)	Fusion protein, function unknown
			T-plastin exon 1	5'reg ss 6/9 bp, 14hb 10.3/0.99	3'reg 14/18 bp 9 sites, 4/6 bp, 9hb 7 nts 10.0/0.99	11mer (24hb) 11mer (30hb) 11mer (28hb)	Fusion protein, function unknown
Interferon-inducible gene (IFI)-54K, 1st coding exon	5'reg 6/9 bp, 14hb 11 nts 11.3/1.00	Sequence information not found	None	Fusion protein, function unknown			
HIV-1 nef	5'cry at aa 66 <sup>b</sup> U1 snRNA 6/9 bp, 14hb ESE 14 nts Sc 12.7/Sc 0.74  5'cry at aa 77 <sup>b</sup> U1 snRNA 6/9 bp, 15hb ESE 14 nts Sc 12.3/Sc 0.78	Not relevant	SV40 T-antigen	3'reg 12/18 bp 6 sites, 6/6 bp, 11hb 18 nts 6.5/0.98	5'reg 5/9 bp, 12hb 10 nts 8.7/0.74	11mer (28hb)	Fusion protein, function unknown
			KIAA1454	3'cry No sequence information 7.3	5'cry No sequence information	10mer (25hb) 11mer (27hb) 11mer (30hb) 11mer (29hb)	Not allocatable
			<i>Cercopithecus aethiops</i> exons 1, 2, 3	No splice junction information	No splice junction information	11mer (26hb)	Not allocatable

two precursors originate from the same transcriptional locus such as in the case of homologous including SV40 RNA *trans*-splicing which occurs between two identical RNA precursors. Further, splicing *in trans* is strongly enhanced between RNA precursors that harbour complementary antisense binding domains which can trigger intron-like bridging between splice sites [4,5]. For homologous SV40 RNA *trans*-splicing the following *trans*-splice helper functions were reported: 1. the early SV40 transcript was found to harbour two binding domains (BDs) of 13 and 10 nucleotides (nts) in length respectively. 2. In addition, SV40 RNA *trans*-splicing was enhanced by mutational weakening of the T-ag 5' donor splice site (ss) which impaired *cis*-splicing in favour of the alternative *trans*-splice reaction towards 3. a strong cryptic 5'ss within a second pre-mRNA molecule. In consequence of the T-ag 5'ss mutation, only 4 instead of 5 nts of the 9 nt 5' splice donor can base-pair to the U1 snRNA of the spliceosomal U1 snRNP, likely impeding its recruitment to this 5'ss. The alternatively used cryptic 5'ss can bind with 6 of 9 nts to the U1 snRNA and hence, represents a much stronger splice donor site. 4. Finally, the alternative cryptic 5'ss was further strengthened by an upstream purine-rich exonic splice enhancer (ESE). SV40 *trans*-splicing was confirmed by the systematic exclusion of alternative mechanisms that might trigger the generation of the observed chimeric RNA, such as direct coding from integrated rearranged SV40 DNA and alternative RNA *cis*-splicing of a long pre-mRNA transcript that might originate from genomic SV40 tandem integration, using southern hybridization and PCR. The coincidence of various *trans*-splicing helper functions led us to the conclusion that SV40 RNA *trans*-splicing is certainly not a chance event but instead a mechanism that is intrinsic to the virus leading to the diversification of viral gene expression [6].

We used computational and manual analyses to screen the adenovirus type 5 exon 2 and the HIV-1 nef sequences as well as the reported *trans*-splicing partner sequences for *trans*-splicing helper functions deduced from the SV40 system. The adenoviral pre-mRNA was reported to splice *in trans* with various cellular host pre-mRNAs including the v-abl Abelson murine leukaemia virus protooncogene homolog 2 transcript variant a, zinc finger protein 41 exon 3, armadillo protein P0071 5'UTR, GTP-binding protein (RAB5) 1st coding exon, GAPDH exons 3 to 7,  $\beta$ -actin exon 3, T-plastin exon 1, and 1st coding exon of the interferon-inducible gene (IFI)-54K during the natural course of infection in human (U521MG, JEG3) cell lines [3]. Adenovirus type 5 is considered to be a non-integrating virus which excludes that the chimeric RNAs are directly DNA coded or generated by alternative RNA *cis*-splicing; thus the described fusion RNAs are most likely the product of a *trans*-splice process. The HIV-1 nef mRNA was reported to splice *in trans* with the SV40 pre-mRNA, KIAA1454, and *Cercopithecus aethiops* genes 1, 2, and 3 in the cell lines CV1 and Cos7 [2]. The setup in these HIV-1 experiments did not resemble the situation of a natural infection and the HIV-1 nef sequence was either stably integrated into the host cell genome to monitor *trans*-splicing with cellular targets or alternatively microinjected as *in vitro* transcribed RNA together with an SV40 transcript. Thus, whilst hybrid RNAs formed with cellular sequences could originate from mechanisms other than *trans*-splicing, the chimeric RNAs formed between HIV-1 nef and SV40 could only be generated by RNA *trans*-splicing.

We (i) screened the above mentioned sequences for the presence of potential antisense BDs, (ii) investigated the strength of the splice donor and acceptor sites involved in RNA *trans*-splicing and (iii) alternative

*cis*-splicing options available to the respective sequences, and (iv) searched for putative ESEs supporting either the *trans*-splice or regular *cis*-splice sites (Table 1).

As for the BDs we considered complementary sites of ten or more consecutive nucleotides being suitable for the fixation of two encountering RNAs under physiological conditions. For all but the adenovirus::IFI-54K sequence pair we identified between one and nine BD candidates. Each pair can recruit at least one 11mer BD with the longest interacting domains being 15 and 16 nts in length (adenovirus::GAPDH). Though the identified BDs can support the *trans*-splice reactions, many complementary BDs of 12 nts or shorter occur with a probability close to one among sequences with a considerable length. The expected number  $N_{\text{exp}}(L_{\text{BD}})$  of BDs of a given length  $L_{\text{BD}}$  (nt) is in good approximation given by Eq. (1)

$$N_{\text{exp}}(L_{\text{BD}}) = L_A L_B / 4^{L_{\text{BD}}} \quad (1)$$

with  $L_A$  and  $L_B$  as the respective lengths of the *trans*-splicing RNAs A and B. The probability  $P$  to observe a BD of the length  $L_{\text{BD}}$  is then given by

$$P = N_{\text{exp}}(L_{\text{BD}}) / N_{\text{obs}}(L_{\text{BD}}) \quad (2)$$

with  $N_{\text{obs}}(L_{\text{BD}})$  as the actually observed frequency of BDs with the length  $L_{\text{BD}}$ . Accordingly, the majority if not all of viral pre-mRNAs have the potential to bind towards mammalian including human pre-mRNAs thus harbouring an important albeit not sufficient molecular feature for splicing *in trans*. The identified potential interactions between viral sequences and their targets as listed in Table 1 are summarised in Table 2.

The chance that two nuclear RNAs encounter and bind to each other via BDs directly correlates with their nuclear abundance and high RNA abundance has to be considered an essential condition for *trans*-splicing. Hence, it was not surprising that Kikumori and colleagues [3] detected *trans*-splicing of adenoviral RNA with the highly abundant cellular pre-mRNAs of  $\beta$ -actin and GAPDH when searching for it.

To investigate the strength of splice donor sites we considered the number of consecutive base pairs and hydrogen bonds (hb) that can be formed between the respective donors with the spliceosomal U1 snRNA. A very strong donor can form up to 9 bp with the U1 snRNA including canonical Watson–Crick as well as wobble base pairs. To assess the strength of splice acceptor sites we counted matches, base pairs, and hydrogen bonds between the U2 snRNA and the branch point consensus sequence as well as sequence matches with the poly U/C stretch consensus sequence which is being recognised by the spliceosomal factor U2AF65. In order to additionally consider the base sequence, i.e. its compliance with the splice donor consensus sequence, we used the algorithms Alternative Splice Site Predictor (ASSP) and Splice Site Prediction (SSP) NNSPLICE 0.9 version to predict the strength of each splice site reflected by a numeric score [7,8]. Accordingly, ASSP-scores of 10 and above and SSP-scores close or equal to 1 have to be considered high. Finally, we searched for purine (A/G)-rich sequences of at least 7 nts or longer that might act as ESEs [9]. The molecular features of the SV40 *trans*-splice reaction are included in the upper part of Table 1 for reference. The involved strong cryptic splice donor has no suitable downstream acceptor and splices *in trans* with the regular SV40 splice acceptor site. Whilst the wild-type T-ag 5'ss efficiently

#### Notes to Table 1:

<sup>a</sup>*Trans*-splicing is enhanced with A to C mutation in 5'ss of exon 1. <sup>b</sup>Cry 5'ss at aa 66 *trans*-splices with SV40, KIAA1454, and *C. aethiops*; cry 5'ss at aa 77 *trans*-splices with SV40 and *C. aethiops*.

ss = splice site, alt = alternative, reg = regular, cry = cryptic, hb = hydrogen bonds, and aa = amino acid.

U1 snRNA recognizes the 5'ss, consensus GUCCAUUCAUA...m3G; RNA-Match: YRGGURRGU; GU = 100% invariant

U2 snRNA recognizes the branch point sequence, consensus AUGAU\_G; RNA-Match: URYURAY; UxA = 100% invariant, A: branch point

U2AF65 recognizes the poly U/C stretch, consensus UUUUUU (U/C) CC(C/U) UUUUUUCC

ESE: purine A/G rich exonic splice enhancers (not indicated if not relevant)

Score predicted by the ASSP software (not indicated if not predicted)

Score predicted by the BDGP SSP software (not indicated if not predicted)

**Table 2**  
Possible base pairing between viral sequences and target RNAs.

Viral transcript	Trans-splicing partner	Potential binding domains		
		5'-virus-3'	3'-ts-partner-5'	
SV40 exon 2	SV40 exon 2	10mer	5'-AGCUUUAAAU-3' 3'-UCGAAAUUUA-5'	
		13mer	5'-GUGGUUUUGUCCAA-3' 3'-CAUCAAACAGGUU-5'	
Adeno-virus type 5 exon 2 of major late transcript (MLT)	v-abl Abelson murine leukaemia viral proto-oncogene homolog 2 transcript variant a (exon 1A)	12mer	5'-GGAUAAAUGGAG-3' 3'-CCUAAUUUACCUC-5'	
		11mer	5'-UUUUCCACAGC-3' 3'-AAAAGGUGUCG-5'	
	12mer	5'-CUGCAGCCAGGG-3' 3'-GACGUCGGUCCC-3'		
	Armadillo protein P0071 5 UTR	12mer	5'-GAGCUGGUGCUG-3' 3'-CUCGACCACGAC-5'	
		12mer	5'-CCAAAAACGAGG-3' 3'-GGUUUUUUGCUCC-5'	
		13mer	5'-GUCAUCCUGAGC-3' 3'-CAGUAGGGACUCG-5'	
		13mer	5'-GGUAUCCACCAA-3' 3'-CCAUAAGGUGGUU-5'	
		GTP-binding protein (RAB5) 1st coding exon	11mer	5'-GUCAUCAAGAC-3' 3'-CAGUAGUUCUG-5'
			11mer	5'-AGGACUUGCUU-3' 3'-UCCUGAACGAA-5'
			11mer	5'-GCUAGGUCGGC-3' 3'-CGAUCACGCGC-5'
			11mer	5'-ACUCCUCCUC-3' 3'-UGAGGAGGAGG-5'
			12mer	5'-GGCAUUUUUUA-3' 3'-CCGUUAAAAAU-5'
			GAPDH exons 3, 5, 7	12mer
	12mer			5'-AGGAGACUUUG-3' 3'-UCCUCCUGAAC-5'
	12mer			5'-GGAGCAGAGAG-3' 3'-CCUCGUCUCUG-5'
	12mer	5'-GGAUCCCUUGA-3' 3'-CCUACGGGAACU-5'		
	13mer	5'-AGGCGCAGCAUCC-3' 3'-UCCGCGUCGUAGG-5'		
	13mer	5'-GGAGCGCAGGUU-3' 3'-CCUCCGUCCAA-5'		
	13mer	5'-AGCUUCCGUUCA-3' 3'-UCGAAGGGCAAGU-5'		
	15mer	5'-AAAGAGGGAGCGGG-3' 3'-UUUCUCCUCGCC-5'		
	16mer	5'-GGCGCUGCAGGAAGG-3' 3'-CCGCGAGUCUUCC-5'		
	12mer	5'-CGGCGUGGUGG-3' 3'-GCCGCCACCACC-5'		
	13mer	5'-UCCCGCGGCGCG-3' 3'-AGGGCGCCGCGC-5'		
	13mer	5'-UGAGCUGCGAGAA-3' 3'-ACUCGACGCUCUU-5'		
	T-plastin exon 1	11mer	5'-AACAAAAGAU-3' 3'-UUGUUUUUCUA-5'	
		11mer	5'-CUCAGCUCCCC-3' 3'-GAGUCGAGGGG-5'	
		11mer	5'-GCCUUUGGAA-3' 3'-CGGAAACCCUU-5'	
		-	-	
		-	-	
	HIV-1 nef	Interferon-inducible gene (IFI)-54K, 1st coding exon	-	-
		SV40 T-antigen	11mer	5'-AAGCAGCUGCU/3' 3'-UUCGUCGACGA-5'
		KIAA1454	10mer	5'-CCACUGCUUA-3' 3'-GGUGACGAAU-5'
11mer			5'-UUCAUCACGUG-3' 3'-AAGUAGUGCAC-5'	
11mer			5'-UGGGGUGGAG-3' 3'-ACCCACCCUC-5'	
11mer			5'-GUGAGCCUGCA-3' 3'-CACUCGACGU-5'	
11mer			5'-GAAGUGUUAGA-3' 3'-CUUCACAAUCU-5'	
C. aethiops exons 1, 2, 3	11mer	5'-GAAGUGUUAGA-3' 3'-CUUCACAAUCU-5'		

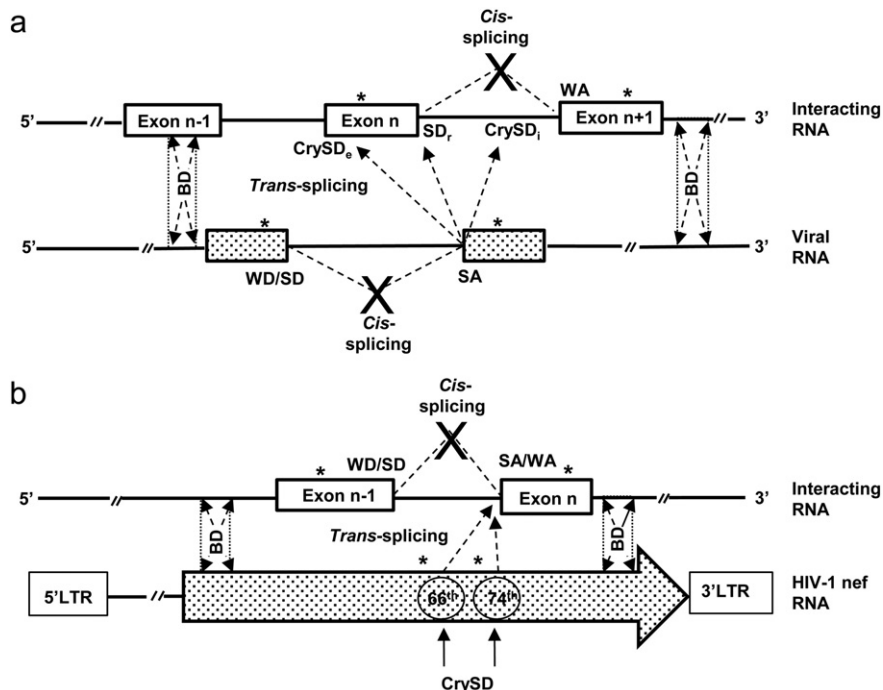
competes with the cryptic splice donor for *cis*-splicing, *trans*-splicing is strongly enhanced by mutational weakening of the T-ag 5' donor ss. In the case of the adenoviral transcript, a strong regular viral splice acceptor *trans*-splices with a variety of mostly strong splice donor sites within cellular pre-mRNAs. The corresponding regular cellular *cis*-splice acceptor sites are mostly weak and the viral *cis*-splice donor site that belongs to the *trans*-splicing acceptor was assigned a lower score as well. In the case of HIV-1 nef two alternative strong cryptic splice donors were described to *trans*-splice with either SV40 RNA or two cellular mRNAs. These cryptic donor sites have no corresponding HIV-1 nef acceptor sites and the recruited SV40 acceptor represents a rather strong ss.

In summary, our data point towards a general pattern associated with viral RNA *trans*-splicing which is schematically depicted in Fig. 1. Accordingly, a strong viral cryptic or regular donor or acceptor ss splices *in trans* with a rather strong regular or cryptic acceptor or donor site. Notably, splice sites that could compete with the *trans*-splicing sites by *cis*-splicing within the respective transcripts mostly are weaker or inexistent. As expected, the majority of splice sites involved in *trans*-splicing are supported by ESEs which can be identified in proximity of many mammalian splice sites.

To experimentally investigate our hypothesis of ss selection/preference in RNA *trans*-splicing we set up an artificial test system comprising a mini-gene, composed of three exonic segments (E1 to E3) joint by two introns (I1 and I2), and two *trans*-splicing RNAs, one for 5' and one for 3' exon replacement harbouring exons E1\* or E3\* respectively (Fig. 2). Whilst I1 harbours a very strong donor (D1) and an acceptor ss (A1) of medium strength, I2 was flanked by strong donor (D2) and strong acceptor (A2). Both *trans*-splicing RNAs were equipped with strong donor (D) or acceptor (A) sites, which can be considered cryptic splice sites since partnering *cis*-splice sites are not available, and BDs guiding the *trans*-splicing sequences to the respective 5' or 3' terminal intron. Consequently, the 5' or 3' splice sites of the *trans*-splicing RNAs could choose either of the available acceptor or donor splice sites, though one would expect that the target splice sites in proximity to

the binding site is preferred [10]. HepG2 cells were co-transfected with plasmids expressing the mini-gene and one of the *trans*-splicing RNAs, total RNA was isolated 24 h post transfection, and *trans*-spliced RNAs were quantified using real-time reverse transcriptase PCR (rtRT-PCR) and the beta-actin mRNA as internal standard. For both *trans*-splicing RNAs, the fold change values indicate that alternative *trans*-splicing towards the more distant target ss was favoured over the expected *trans*-splicing with the respective proximal ss. In detail, D-to-A2 *trans*-splicing between the very strong donor D and the strong acceptor A2 was favoured (2.2-fold) over D-to-A1 *trans*-splicing from D to the weak acceptor A1; similarly, D1-to-A *trans*-splicing between the very strong donor D1 and the very strong acceptor A was favoured (7.2-fold) over D2-to-A *trans*-splicing between strong donor D2, which was a weaker donor than D1, and A. Notably, the alternative *cis*-splice partner ss D2 or A1 of the favoured *trans*-splice target sites A2 and D1 were weaker than partner sites D1 or A2 of the disfavoured *trans*-splice target sites A1 or D2, respectively. By qualitative agarose gel electrophoresis and by monitoring the melting profile of the rtRT-PCR products we excluded that longer products such as E1\*/E2/E3 or E1/E2/E3\* were generated and co-detected during the quantification of distal E1\*/E3 or E1/E3\* exon joining. These data support our model of ss selection during RNA *trans*-splicing indicating that (i) stronger target splice sites are favoured over weaker sites (ii) in particular when the alternative *cis*-splice partner site of the target is weaker compared with the ss of the *trans*-splicing RNA.

The chimeric RNAs generated in the course of heterologous viral RNA *trans*-splicing either (1) did not lead to the translation of fusion proteins, because (1a) multiple stop codons will only allow the formation of a truncated cellular protein such as in the case of the adenoviral *trans*-splice to the v-abl proto-oncogene homolog or because (1b) the translational start site is lost upon *trans*-splicing as in the case of the Armadillo protein with the adenoviral RNA, or (2) yielded fusion proteins of unknown function. Therefore, the highly transforming 100 kDa SV40 super T antigen remains the only example of a functional



**Fig. 1.** Schematic illustration of a *trans*-splicing reaction between a viral pre-mRNA and an interacting pre-mRNA molecule. a, A strong viral cryptic or regular 3' acceptor ss *trans*-splices with a rather strong regular splice donor (SD<sub>i</sub>), cryptic exonic splice donor (SD<sub>e</sub>) or cryptic intronic splice donor (SD<sub>i</sub>). Ss that could compete with the *trans*-splicing sites for *cis*-splicing mostly are weaker or inexistent. b, In the case of the HIV-nef RNA one of two possible strong cryptic splice donor sites (at aa positions 66 and 74) *trans*-splice with a splice acceptor site of a cellular target. a and b: One or more potential binding domains (BDs) of 11 nts or longer can support the *trans*-splice reaction. Notably, in order to allow antisense binding via the BDs one of the two strands has to partly turn around as indicated by the crossing arrows. WD: weak donor; SD: strong donor; WA: weak acceptor; SA: strong acceptor; Cry: cryptic ss; \*: ESE: exonic splice enhancer.



Sequence Detection System (Applied Biosystems, Foster City, CA) using primers F1 5'-AGTCTTCAGGGTGTGTTAGAAAACCA-3', F2 5'-CATCTTGCTGCAAAGCTGAAAA-3', F1\* 5'-CAGGCTCCATACCGACGATC-3', R1 5'-TTCTCATGGCAAAGTCTTCCA-3', R2 5'-AAGCTGCTTTCTCTTAATTCTTTGT AAC-3', and R1\* 5'-TTGCTGTGTCCCGTGATC-3' (AITbiotech, Singapore). Relative *trans*-splice activities were calculated by the comparative Ct method ( $\Delta\Delta Ct$ ) using  $\beta$ -actin as internal control. The indicated fold change values were calculated as  $2^{-\Delta\Delta Ct}$ .

#### Conflict of interest statement

The authors declare that they have no competing financial interests.

#### Acknowledgments

This work was supported by NUS-Cambridge Start-up grant R-182-000-163-646 of the National University of Singapore, New Investigator Grant NMRC/NIG/1058/2011 of the National Medical Research Council of Singapore, and Academic Research Fund Grant R-182-000-200-112 of the Ministry of Education – Singapore, all to V.P.

#### References

- [1] Eul J, Patzel V. Homologous SV40 RNA *trans*-splicing: a new mechanism for diversification of viral sequences and phenotypes. *RNA Biol* 2013;10:1–11.
- [2] Caudevilla C, Da Silva-Azevedo L, Berg B, Guhl E, Graessmann M, Graessmann A. Heterologous HIV-nef mRNA *trans*-splicing: a new principle how mammalian cells generate hybrid mRNA and protein molecules. *FEBS Lett* 2001;507:269–79.
- [3] Kikumori T, Cote GJ, Gagel RF. Naturally occurring heterologous *trans*-splicing of adenovirus RNA with host cellular transcripts during infection. *FEBS Lett* 2002;522:41–6.
- [4] Konarska MM, Padgett RA, Sharp PA. *Trans* splicing of mRNA precursors *in vitro*. *Cell* 1985;42:165–71.
- [5] Solnick D. *Trans* splicing of mRNA precursors. *Cell* 1985;42:157–64.
- [6] Patzel V. Viral RNA *trans*-splicing: chance event or product of evolution? *SOJ Microbiol Infect Dis* 2013;1:1–2.
- [7] Wang M, Marín A. Characterization and prediction of alternative splice sites. *Gene* 2006;366:219–27.
- [8] Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4–3:311–23.
- [9] Zheng ZM. Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J Biomed Sci* 2004;11:278–94.
- [10] Mansfield SG, Clark RH, Puttaraju M, Kole J, Cohn JA, Mitchell LG, et al. 5' exon replacement and repair by spliceosome-mediated RNA *trans*-splicing. *RNA* 2003;9:1290–7.
- [11] Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, et al. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res* 2012;40:10345–55.