Virology 456-457 (2014) 329–341

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

A new look at adenovirus splicing

Hongxing Zhao *[,1], Maoshan Chen [1,2], Ulf Pettersson

Department of Immunology, Genetics and Immunology, Uppsala University, S-751 85 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 20 December 2013
Returned to author for revisions
23 January 2014
Accepted 3 April 2014

Keywords:
Adenovirus type 2 infection
RNA splicing
Polyadenylation
cDNA sequencing

ABSTRACT

Adenovirus type 2 RNA splicing events were quantitatively mapped by using deep cDNA sequencing. The majority of the previously identified splice sites were detected. The lack of complete consistency between the present and previous results is because of some sites which were incorrectly mapped in previous studies, such as the splice sites for pVII, pVIII and E3-11.6K. Several previously predicted splice sites such as that for E3-14.5K and E4ORF3/4 were not detected. In addition, several new splice sites were identified. The novel RNAs may code for hitherto undetected proteins or alternatively spliced mRNAs for known proteins. The open reading frames downstream of two novel splice sites, located in the major late transcription unit region, were shown to be highly conserved. Another interesting possibility is that some of them are non-coding RNAs. Finally, the adenovirus mRNA polyadenylation sites were accurately mapped and in some cases shown to be heterogeneous.

Introduction

Adenoviruses are non-enveloped, icosahedral viruses containing a linear, double-stranded DNA molecule (Green et al., 1967). Both strands are transcribed with genes located on both the so-called rightward reading strand (r-strand) and leftward reading strand (l-strand). The genome consists of five early transcription units (E1A, E1B, E2, E3 and E4), two intermediate units (IX and IVa2) and one major late unit that generates five families of late mRNAs (for review (Shenk, 1996)). An additional late l-strand transcription unit encoding the U exon protein (UXP) has been identified recently (Tollefson et al., 2007; Ying et al., 2010). Each transcription unit contains its own promoter, and most of them encode more than two mRNAs by differential splicing of a single linear transcript. Expression of adenovirus genes is regulated during the productive infection in a step-wise manner. The immediate early gene E1a is expressed first followed by the expression of the delayed early genes, E1B, E2, E3 and E4. Then the intermediate early genes, IVa2 and IX are expressed. L1, a late transcript from the major late promoter, is also made during this phase. Soon after the onset of viral DNA synthesis, the transcription is switched from an early to a late mode for the production of viral structural proteins. Adenovirus makes an extensive use of alternative RNA splicing to produce a very complex set of mRNAs. Except for the polypeptide IX mRNA, all adenovirus primary transcripts undergo one or more splicing events which give rise to about fifty distinct mRNAs during a lytic infection. From early region E1A, three major mRNAs, the 13S, 12S and 9S mRNAs, and two minor mRNAs, the 10S and 11S mRNAs, are produced by alternative splicing (Berk and Sharp, 1978; Chow et al., 1979; Perricaudet et al., 1979; Schmitt et al., 1987; Spector et al., 1978; Virtanen and Pettersson, 1983). These mRNAs have common 5′ and 3′ ends, but differ from each other by the size of the intron. The E1A 13S and 12S mRNAs are the most abundant RNA species early after infection, while the 9S mRNA represents less than 5% of the total E1A mRNAs. At late times, a shift in the steady-state levels of the mRNAs occurs and the 9S mRNA becomes the most abundant species. Transcription of region E1B generates two major mRNAs (22S and 13S) and two minor mRNAs (14.5S and 14S) by splicing of one or two introns from a common precursor RNA (Berk and Sharp, 1978; Spector et al., 1978; Virtanen and Pettersson, 1983). The 22S and 13S mRNAs are the predominant species whereas the 14.5S and 14S mRNA represent less than 5% of the total steady-state level of E1B mRNAs (Babich and Nevins, 1981). Following the progression of the infection, the abundances of E1B 22S and 13S mRNAs are changed from equal amounts during the early stage to approximately 20-fold excess of 13S mRNA at late times. Region E2 is transcribed from the l-strand by using alternative promoter sites for the initiation of transcription. Two major classes of transcripts, E2A and E2B, have been described (Chow et al., 1979; Stillman et al., 1981). Although multiple mRNAs are produced from E2A, one mRNA encoding a single-stranded DNA-binding protein (DBP) is dominant. It is produced by the removal of two introns from the pre-mRNA. E2B precursor RNAs bypass the polyadenylation signal used for E2A mRNAs and extend to a downstream polyadenylation site. Two major proteins, the 87K

* Corresponding author. Fax: +46 18 471 4808.
E-mail address: Hongxing.Zhao@genpat.uu.se (H. Zhao).
[1] These authors contributed equally to this work.
[2] Present address: Department of Biochemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3083, Australia.

http://dx.doi.org/10.1016/j.virol.2014.04.006
0042-6822/

terminal protein (pTP) and the 140K DNA polymerase (Adv-Pol), are translated from the E2B mRNAs. They are both involved in adenovirus DNA replication (Smart and Stillman, 1982; Stillman et al., 1981, 1982). During the late stages of infection UXP is expressed abundantly from the l-strand from its own promoter, but its function remains unknown (Tollefson et al., 2007; Ying et al., 2010). The E3 transcription unit is embedded within the major late transcription unit and several of the splice sites are utilized in the maturation of the fiber mRNA. Previous studies have identified at least nine E3 mRNAs which are generated by differential processing of two major RNA precursors, E3A and E3B. The precursors share a common cap site, but differ from each other by having different 3′-ends. The E3B has one poly(A) addition site at 30846, whereas E3A has four poly(A) addition sites at 29792, 29799, 29801 and 29804. The E3 mRNAs are translated into seven proteins, namely 12.5K, 6.7K, gp19K, 11.6K, 10.4K, 14.5K and 14.7K (Wold et al., 1995). These proteins play a very important role in counteracting the antiviral defenses of the host (Gooding, 1992; Gooding and Wold, 1990). The E4 transcription unit is located at the right end of the l-strand (Pettersson et al., 1976; Sharp et al., 1974). A single primary transcript is spliced and generates a very complicated set of mRNAs and as many as 24 mRNAs with identical 5′ and 3′ ends have been reported (Freyer et al., 1984; Herisse et al., 1981; Rigolet and Galibert, 1984; Tigges and Raskas, 1984; Virtanen et al., 1984). E4 contains seven open reading frames (ORFs), and all except ORF3/4 polypeptides have been detected in infected cells.

Nearly all of transcription from the major late transcription unit (MLTU) starts at the onset of adenovirus DNA replication. Primary transcripts are generated from a single promoter and they are polyadenylated at one of five different sites, generating five families of RNAs, L1–L5. A minimum of 20 mRNAs are generated from the MLTU, and nearly all of them have a common 201-nucleotide tripartite leader sequence at their 5′ ends. Some of them contain an additional 440-nucleotide long i-leader exon. The L1 pre-mRNA is transcribed both at intermediate and late times, but its splicing pattern changes with time. The mRNA which codes for the two structurally related polypeptides 52K and 55K is produced both at intermediate and late times after infection, whereas the IIIa mRNA is produced exclusively in the late phase. Two poly(A) addition sites at 14113 (the most abundant) and 14119 were identified (Hales et al., 1988; Prescott and Falck-Pedersen, 1994).The L4-100K and L4-22K mRNAs are transcribed at a relatively early time of the late phase (Larsson et al., 1992). The L4-22K mRNA is transcribed from an internal L4 promoter embedded in the ORF of L4-100K (Morris et al., 2010). The L4-22K protein suppresses adenovirus early gene expression, but activates the full panel of L1-L5 transcription. The L4-33K protein is virus-encoded alternative RNA splicing factor and has been shown to activate splicing of Ad late gene transcripts with weak 3′ splice sites (Tormanen et al., 2006). Furthermore, L4-33K protein regulates selective accumulation of Ad late gene transcripts (Wu et al., 2013). All remaining late mRNAs are produced in the late phase. The L2 pre-mRNA is spliced into four major mRNAs with a common poly(A) addition site at nt 17969 and encoding polypeptides pIII (penton base), pV (major core protein), pVII (core protein) and pX (pμ) (Akusjarvi and Persson, 1981; Le Moullec et al., 1983); the L3 pre-mRNA is spliced into three major mRNAs with a common poly(A) site at nt 22443 which encode polypeptides pVI (hexon associated protein), pII (hexon) and the 23K viral protease (Prescott and Falck-Pedersen, 1994); the L4 pre-mRNA is spliced into four polypeptides, 100K, 22K, 33K and pVIII (hexon associated protein) with a common poly(A) site at nts 28223 and 28228 (Sittler et al., 1994); and finally, L5 includes transcripts encoding only pIV (fiber) with a poly(A) addition site at nt 32798. L5 mRNAs comprise a family of transcripts with a number of different 5′-leader sequences (x, y, z leaders) in various combinations in addition to leaders 1, 2, 3 and i (Le Moullec et al., 1983; Uhlen et al., 1982).

For decades, adenoviruses have served as an outstanding model system to study the molecular mechanisms of splicing due to the simplicity of their genomes and their efficient mode of replication. Most adenovirus mRNAs are generated by the removal of one or more introns and most of these introns are located in the 5′ or 3′ noncoding portion of pre-mRNA. Thus the viral introns do only in a few cases interrupt the ORFs. The development of high throughput sequencing methods has facilitated the discovery of many novel transcribed regions and splicing isoforms (Djebali et al., 2012). It is also a very powerful tool to study alternative splicing under different conditions at an unprecedented depth. Here we present a comprehensive analysis of adenovirus RNA splicing during different phases of the infection and a complete adenovirus splicing map. Two deep sequencing experiments, single-end and paired-end sequencing, were performed. Single-end sequencing was done by using the standard single-read DNA library preparation. The major shortcoming of this procedure is the short sequence reads and an exponential increase in error rates along the reads (Cox et al., 2010). The more recently developed paired-end sequencing allows for reading 255 nt long sequences from both ends of cDNA fragments. The data generated from paired-end sequencing, utilized in our second experiment, should thus be more reliable.

Results and discussion

Summary of sequencing results

Using mRNA single-end and total RNA paired-end sequencing technologies, the Ad2 RNA splicing profile during different phases was studied. Infection and RNA isolation were performed as in our previous study (Zhao et al., 2007). Briefly, synchronized human primary lung fibroblasts (IMR-90) were infected at a multiplicity of 100 FFU/cell. Infected cells were collected at 6, 12, 24, and 36 hpi. As shown in our previous study, these time points represent different stages of the infectious cycle, i.e. before any adenoviral gene expression, after immediate early gene (E1a) expression, after the onset of adenoviral DNA replication, and after late gene expression, respectively (Zhao et al., 2007). Thus, we could correlate the expression of RNAs with the progression of the infection. Single- and paired-end sequencing was performed. Only three RNA samples from 12 hpi, 24 hpi and mock were subjected to single-end sequencing in a pilot experiment whereas all RNA samples including 6 hpi, 12 hpi, 24 hpi, 36 hpi and mock were sequenced by paired-end sequencing. Single-end sequencing yielded 50–54 million 76 bp long sequence reads per sample, as shown in our previous publication (Zhao et al., 2012). The fraction of reads that aligned to the adenovirus genome increased dramatically, from 1.3 million reads at 12 hpi to 15.9 million reads at 24 hpi, indicating a very efficient infection (Table 1). The 401 reads that were aligned to the adenovirus genome in the mock sample represented the background noise. In the case of paired-end sequencing 30 million of 255 bp long sequence reads per sample were generated. The sequence reads that mapped to the adenovirus genome increased dramatically after 12 hpi. By using TopHat (Trapnell et al., 2009), an efficient read-mapping algorithm designed to align the sequence reads to reference genome without relying on known splice sites, 2228 and 1460 adenovirus splice junctions (with more than 1 sequence reads) were identified by single-end and paired-end sequencing, respectively. Although more splice junctions were identified by single-end sequencing, their sequence coverage was much lower than seen with paired-end sequencing. In general, the number of sequence reads