Virology 511 (2017) 344–353

Contents lists available at ScienceDirect

Virology

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

Hairpin structures with conserved sequence motifs determine the 3′ ends of non-polyadenylated invertebrate iridovirus transcripts

İkbal Agah İnce[a,*], Gorben P. Pijlman[b], Just M. Vlak[b], Monique M. van Oers[b]

[a] Department of Medical Microbiology, Acıbadem University Medical School, Atasehir, 34752 Istanbul, Turkey
[b] Laboratory of Virology, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

ARTICLE INFO

Keywords:
Invertebrate iridescent virus
Chilo iridescent virus
Iridovirus
RNA processing
mRNA termination
Non-polyadenylated transcripts
LACE
Hairpin structures
Inverted repeats

ABSTRACT

Previously, we observed that the transcripts of *Invertebrate iridescent virus 6* (IIV6) are not polyadenylated, in line with the absence of canonical poly(A) motifs (AATAAA) downstream of the open reading frames (ORFs) in the genome. Here, we determined the 3′ ends of the transcripts of fifty-four IIV6 virion protein genes in infected Drosophila Schneider 2 (S2) cells. By using ligation-based amplification of cDNA ends (LACE) it was shown that the IIV6 mRNAs often ended with a CAUUA motif. *In silico* analysis showed that the 3′-untranslated regions of IIV6 genes have the ability to form hairpin structures (22–56 nt in length) and that for about half of all IIV6 genes these 3′ sequences contained complementary TAATG and CATTA motifs. We also show that a hairpin in the 3′ flanking region with conserved sequence motifs is a conserved feature in invertebrate-infecting iridoviruses (genus *Iridovirus and Chloriridovirus).*

1. Introduction

Eukaryotic cells use three RNA polymerases to transcribe genes, each with specific mechanisms of initiation, elongation and termination (Richard and Manley, 2009). Precursor protein-encoding transcripts (pre-mRNAs and/or hnRNAs) are produced by RNA polymerase II (RNAP II). These pre-mRNAs undergo several processing events before they reach their mature mRNA form (Gromek and Dvir, 2010; Moore and Proudfoot, 2009). The first step is the co-transcriptional addition of a methylated-guanosine cap to the 5′-end of the transcript. In the next step, introns may be removed by splicing and the growing transcript will be packaged with RNA-binding proteins. Finally, the majority of eukaryotic RNAP II transcripts are cleaved at a predetermined site and become polyadenylated at their 3′-end. The resulting poly(A) tail plays an important role in nucleo-cytoplasmic export and is important for mRNA stability and translation efficiency (Moore and Proudfoot, 2009).

The processing of the 3′ end of most eukaryotic pre-mRNAs is dependent on a canonical polyadenylation signal, present downstream of the translational stop codon and found upstream of the cleavage and polyadenylation site. This signal is represented by the hexanucleotide AAUAAA in 80% of the mammalian pre-mRNAs and slight variants thereof, may also be used (see e.g. Birnstiel et al., 1985; Edwalds-Gilbert et al., 1997). The AAUAAA motif is often accompanied by a U- or GU-rich element located downstream of the cleavage site (Chen and Wilusz, 1998; McLauchlan et al., 1985). These motifs are recognized by the cleavage and polyadenylation specificity factor (CPSF) (see e.g. the review by Richard and Manley (2009) and Elkon et al. (2013)), which cleaves typically between 10 and 30 nucleotides (nt) downstream of the poly(A) motif (Sachs and Wahle, 1993). After the cleavage, a stretch of 200–300 adenines is added by the enzyme polyadenylation polymerase. Poly(A)-binding protein (PABP) binds to this poly(A) tail to protect the new mRNA from exonuclease activity. The poly(A)tail also plays an important role in translation re-initiation, which also involves PABP (Eliseeva et al., 2013; Moore and Proudfoot, 2009).

Previously we observed non-polyadenylated mRNAs for *Insect iridescent virus 6* (IIV6) in the genus *Iridovirus*, family *Iridoviridae* (Nalçacioğlu et al., 2003). In line with this observation, the 3′-untranslated regions of the 211 predicted IIV6 open reading frames (ORFs) (Eaton et al., 2007) do not contain canonical eukaryotic polyadenylation signals (AATAAA) or close variants thereof. The question thus is, how the 3′-ends of IIV6 transcripts are determined and whether they use another strategy to prevent exonuclease activity.

Non-polyadenylated mRNAs are rare in eukaryotes. The best-known examples are the metazoan histone transcripts synthesized in the S-phase of the cell cycle (Marzluff et al., 2008). The mature histone mRNAs have a conserved sequence 25–26 nt in length, consisting of a stem–loop structure with a 6 nt base stem and a 4 nt loop, followed by 5 additional nucleotides before the transcript ends (Marzluff et al., 2008). The histone downstream element (HDE), present in the pre-

* Corresponding author at: Department of Medical Microbiology, Acıbadem University Medical School, Atasehir, 34752 Istanbul, Turkey.
E-mail addresses: ikbal.agah.ince@gmail.com, agah.ince@acibadem.edu.tr (İ.A. İnce).

http://dx.doi.org/10.1016/j.virol.2017.06.026
Received 1 April 2017; Received in revised form 21 June 2017; Accepted 22 June 2017
Available online 11 July 2017
0042-6822/

mRNA is required for 3′ end cleavage and associates with U7 small nuclear (sn) RNA via base pairing (Marzluff et al., 2008). The CPSF subunit 73 is responsible for histone pre-mRNA cleavage (Dominski et al., 2005; Mandel et al., 2006). The stem-loop structure interacts with the stem–loop-binding protein (SLBP) to form a complex, thereby stabilizing the mRNA (Marzluff et al., 2008; Richard and Manley, 2009). Several authors also reported alternative, non-canonical signals to determine mRNA 3′-ends (Wilusz et al., 2008; Wilusz and Spector, 2010; Yang et al., 2011). In that case, the pre-mRNA processing was mediated by enzymes normally involved in tRNA biogenesis or pre-mRNA splicing.

In the current study, we used a universally applicable RT-PCR method (Ligation-based amplification of cDNA ends, LACE) that we have described previously (İnce et al., 2013) to determine the 3′-ends of non-polyadenylated transcripts This method was applied to the transcripts of the 54 previously determined IIV6 virion protein genes (İnce et al., 2010). In addition, we analysed the 3′-flanking regions of all IIV6 ORFs *in silico* (as well as those of the related viruses IIV3, IIV 9 and IIV31) for sequence motifs that are more frequently present in these areas than would be expected from the overall nucleotide composition of the genome in order to identify conserved motifs that may play a role in transcript 3′ end formation.

2. Results

2.1. Many IIV6 virion protein transcripts terminate after a CAUUA motif

To determine the 3′-end of IIV6 transcripts, we recently developed a method, called LACE that can be applied to non-polyadenylated mRNAs (İnce et al., 2013). In the first step, a phosphorylated DNA oligonucleotide is added to the 3′-ends of the total pool of transcripts using RNA-ligase. This linker sequence is then used to anneal to a linker-specific primer, which allows cDNA synthesis followed by gene specific PCR-amplification. Using this method, viral transcripts amplified from IIV6 infected Drosophila S2 cells over the course of infection, were analysed. The transcripts from the 54 structural virion proteins of IIV6 (İnce et al., 2010) were selected for this analysis. In this way, we made sure that we only tested functional ORFs. This analysis was successfully performed for 46 out of the 54 analysed ORFs. Remarkably, 26 of the analysed transcripts ended after a CAUUA motif (Table 1; CATTA in cDNA). For ORF 309L and 380R, the transcripts ended after the CAUUA-variants CUUUA and CAUAA, respectively (Table 1; CTTTA and CATAA in cDNA). The transcript of ORF 285L had an additional adenosine at the 3′-end (CAUUAA). These results suggest that the CAUUA motif plays an important role in 3′-end formation in the majority of the IIV6 gene transcripts. The other 17 transcripts for which we were able to experimentally determine the 3′-end had no CAUUA sequence in their 3′-untranslated regions. These transcripts ended after various UA-rich motifs, of which UUAAA was most frequently found (Table 1).

2.2. Enriched CATTA and TAATG motifs downstream of IIV6 ORFs

Next, the IIV6 genome-wide screen was performed to identify sequence motifs that were enriched in the 100 and 200 bp regions downstream of all predicted IIV6 ORFs (Eaton et al., 2007). When searching for 4-mer motifs, the ATTA motif was strongly enriched. The *in silico* analysis of pentanucleotide motifs showed that the CATTA motif was most prominently enriched, appearing 136 times in the 100 nt downstream regions out of the 211 non-overlapping ORFs recognized by Eaton et al. (2007), with a 2.4-fold enrichment as compared to the whole IIV6 genome (Table 2). The complementary TAATG motif was observed 131 times in the 100 nt downstream (with an enrichment factor of 2.3). Various TA-rich motifs also showed up during the analysis with ATTAA and the complementary TAATT motif being most enriched (Table 2). These motifs occur very often in the whole genome as well (2548 times each), which may be in part a reflection of the AT-rich (71.4%) nucleotide composition of the genome. In the experimental analysis 17 out of the 46 transcripts that we analysed ended after AU-rich motifs (see Table 1). When we performed the *in silico* analysis with 3′-flanking regions of 200 nt instead of 100 nt, the CATTA and TAATG motifs were found 200 and 192 times in these downstream regions, respectively, with relative enrichments of 2.8 and 1.7. As we analysed 215 ORFs, and seen the outcome of our experimental analysis, these high numbers may indicate that some ORFs are followed by more than one of these motifs. The ATTAA and TAATT motifs found in the 200 nt may only in part represent functional motifs, as their appearance of over 400 times by far exceeds the number of analysed ORFs (Table 2). When searching for 6-mer motifs downstream of the translational stop codons various CATTA and TAATG containing sequences were found to be enriched (Table 2). The sixth nucleotide showed versatility. For example, when 100 nt long 3′ flanking regions of the tested ORFs were examined, CATTAA and the complementary TTAATG (appearing 102 and 111 times, enriched 4.2 and 3.9 fold, respectively) and CCATTA and the complementary TAATGG (49 and 55 times, with enrichment factors of 3.0 and 3.4, respectively). As the range of conservation did not extend beyond the 5-nt of the CATTA and TAATG motifs, we did not run the analysis for heptanucleotides.

The computational analysis that we performed has the limitation that it does not show whether the CATTA motifs are correlated to individual ORFs or that some regions downstream of ORFs have these motifs more than once. Neither does this method provide information on how often CATTA and TAATG occur together in the same 3′-flanking regions, nor whether they appear in a specific order. The same holds true for the complementary AU-rich motifs. So, this finding does not imply that 131 (or even 192 ORFs when looking at 200 nt) are followed by downstream CATTA and TAATG motifs, but only that a considerable number of ORFs are expected to carry both the CATTA and the complementary TAATG sequence in the 3′-flanking regions. Despite these limitations, the data suggests a tendency for both the CATTA and TAATG motifs to be present together. If both motifs are present in close proximity, a hairpin structure may form in the native transcript (pre-mRNA) and may be also present in the mature mRNA, depending on the position of these two motifs compared to the mRNA 3′ends. We have not carried out a similar analysis for ORFs without CATTA motifs, since it is difficult to predict which AT-rich element will be used in this case.

2.3. Stem-loop structures in 3′ untranslated regions of IIV6 genes

When we looked more closely at the experimentally determined 3′ UTRs (Table 1), we noted that the TAATG sequence was present upstream of the CATTA motif in 24 out of the 26 cDNA sequences ending with CATTA (Table 1, red letters). For ORFs, where the cDNA ended with an AT-rich sequence, no TAATG/CATTA motif was found upstream of the 3′-ends. Inspection of all IIV6 ORFs by eye revealed that 110 of the 215 analysed genes had CATTA motifs in their 3′-regions (Table S1), suggesting that the CAUUA sequence plays an important role in IIV6 mRNA 3′ end determination. The distance of the CATTA motif from the translational stop codon varied from 10 to 133 nt with an average of 40 nt (Fig. 1A). Next we examined the sequence between the stop codon and the CATTA motif and noted that this motif was very often preceded by the complementary TAATG sequence (in 104 out of the 110 ORFs with 3′-flanking regions carrying CATTA (Table S1)). This result is in line with the high enrichment values found for both sequence motifs in the *in*

Table 1
The 3′ends of the IIV6 virion protein transcripts as determined by LACE*. UTR sequences represented as cDNA sequence. The indications L and R refer to the orientation of the gene on the genome (L-being on the complementary strand). Sequences in red represent TAATG motifs, complementary to the CATTA variant at the end. Three examples of co-variance were also observed (blue text/blue boxes).

ORF details			Primer details		3′ UTR characteristics			Temporal class of expression*		
Name	Start position	Stop position	Primer start	Primer sequence	3' UTR determined experimentally	End motif	Position 3'-end	IE	DE	L
010R	2498	2860	2498	atgaataattttaattactttaatg	ND	.....		ND	ND	ND
022L	9277	5762	6242	cctctgtggaatgaatgg	ATTTTAATACCTTTAAGGTATTA	TATTA	5739	x		
034R	11997	12401	11997	atgaaacagaatttattaatcc	TTAAATTTTAATGATAATCTATCATTA	CATTA	12428	x		
061R	27538	28005	27538	atgaacgatagaggtataattat	ND	.....				
084L	35091	34594	35074	atggcaaaagtcattaaa	ND	.....				
096L	39395	38871	39349	gatggtttatatagattttc	TTTAATTTTTTCATTATTTTTTTTATTATTTTTTCATTA	CATTA	38832	x		
104L	43008	42835	42990	atgccacattacgttgttg	ACGGTTTGCGTTTTTAAAGGTCTTTTGATCTTTAAA	TTAAA	42799	x		
111R	43911	44438	43911	atgattttggttttattacttat	ND	.....				
117L	46517	45768	46149	gctatggcattcaaccct	ATTTTAATGGTTTTGTTACCATTA	CATTA	45744		x	
118L	48177	46630	47113	tgggtaaattccgaccca	TTCTTAAAATTTAATGGTAAAAACCATTA	CATTA	46601	x		
123R	49533	49961	49536	atgaacctacgaaaatagtag	AATTTATAATTTTAATGGTTAGATAACCATTA	CATTA	49993	x		
130R	51751	52356	51855	ctattcaggaaatactacattg	ND	.....				
138R	54168	54512	54168	atgaataataatcaacaaatactt	TTTTTAAAAATTTAATGGAAAGTTCCATTA	CATTA	54542	x		
142R	55500	56378	55875	tttattacagctacttatgat	TTGAACTTTTGATTTTAACAGCCAATAGCTGTTAAA	TTAAA	56417	x		
149L	61960	59960	60443	gtgatttctattttacacatg	ATTAAAAATTTTAATGAAATATTTCATTA	CATTA	59931		x	
155L	62770	62003	62356	taccttggatacggaaatg	AAAACATACATAAGAGAGATTCGAAATAAAA	TAAAA	61972	x		
159L	64967	63540	64029	acaagtaaagtaggttctatt	AAAACTTTTAATGAAAAATTTTCATTA	CATTA	63513			x
179R	75567	79127	78625	ttgaattatatatggttgcac	AATATATTGTATATAAAATATATTGTATATAAAATTTTAATGGATATTATCCATTA	CATTA	79183		x	
198R	85637	87055	86553	ttgcaggtcaaacgttaga	TTTATATATTTATTGTAATT	TAATT	87076			x
203L	89334	88852	89313	atgtccattcaaacattaataa	ND	.....				
209R	92277	95417	94917	tgggcgttttgtactggt	ATTTTAATGATTAAAAATCATTA	CATTA	95440	x		
213R	98000	99568	99068	gatagtggatatcatacgtt	TTTTAATGACTATAGTCATTA	CATTA	99590	x		
219L	*102489*	*101458*	*101950*	*gcaggatggacactatct*	TAATAATAACATAATGCCAATGTGAAAATTGTCCATTGTAGAGAAGTGGTGCGCTTAAGTCAGTTCAAGGTATTAACAGCTTTTTAGTTTTAATGAACATTATTCATTA	CATTA	*101325*	*x*		
227L	105597	105406	105578	atgaatcagcttaatcaatt	ATTTTTTTTAATACCTTTTGGTATTAAA	TTAAA	105378	x		
229L	107921	106590	107071	ggagcttttatatcttggcc	GCAAAAGAATAAAGAATGAATTTGAACTATAATT	TAATT	106556		x	
232R	108021	110036	109559	tcactacattgctttagaac	ATAATTTTGTAATTTAATTTTGTAATTTTAATGGAAATATTTCCATTA	CATTA	110084			
234R	110163	110744	110163	agatgtaaaagtaagtgctg	AAATTTAATAGAAACGTTTCTATTAAA	TTAAA	110772			x
261R	121519	125604	125106	taatcccaacacaaccgc	TTTTTTTTTAAATTTTAATGGTTATTAACCATTA	CATTA	125638	x		
268L	127775	125643	126135	gtcgtgatgacaaaataatt	ATTTTAATGGTTAATAACCATTA	CATTA	125620	x		
274L	130158	128755	129248	ctattaaagcacttttcttt	AGGGGGGTTGGGGTTGGTTTTATTATCATTTTTTAATGGAAAAATTCCATTA	CATTA	128702			x
295L	139425	135394	135894	tgtttatggtagacagtaaaa	GTTTTTAATGCCTTCGGCATTAA	CATTAA	135372	x		
307L	142846	142253	142754	aggtataaagggatcttgg	TTTTTAATGGTATTGTACCATTA	CATTA	142230	ND	ND	ND
309L	143800	143147	143536	tcggatgttgataaaggac	CACCATAGATTAAAATATAAAGGTTAGTACCTTTA	CTTTA	143112	x		
312R	143926	144204	143926	atgaaccccgcaattgt	TTAATTTTTAAAAATTTAATGAAATTGTTTCATTA	CATTA	144240	x		
317L	147527	146271	146782	cttatgtctggaacggt	AATTTAATGAAATTGTTTCATTA	CATTA	146248			x
325L	148350	147871	148004	atgtttcttttgagaattttc	TTAACACCTAAAAGGTGTTAAT	TTAAT	147849	ND	ND	ND
329R	148473	149609	149087	tgttattgatggtgttactc	ATTTAAAGGTAATTTACCTTTAA	TTTAA	149633		x	
337L	152755	151517	151971	gaatactgtcttcataatcc	TTTTCTGTTTTAAAGGGTTTAA	TTTAA	151495		x	

(continued on next page)

Table 1 *(continued)*

342R	153531	153797	153531	atggataaacctcgcgaac	ATAATTGAAATATAAAGGTTGTTTAACCTTTATATTTTAAAA	TAAAA	153840			x
355R	158160	158708	158210	atttgaagatataactgact	AATTTGTTGAAACATTA	CATTA	158726	ND	ND	ND
361L	162659	161031	161538	tctcaagttcaaatacagac	ATTTTTAATGACAAAAGTCATTA	CATTA	161008	x		
366R	162769	163113	162769	atgccattattaagaaaacgat	ND	.....			x	
374R	165568	166065	165568	atggatatagaatttggaaat	TTTTAATGGAAATTCCATTA	CATTA	166086		x	
378R	167305	167886	167473	tcatctcctagaaggtc	TCAGTTAACAAAACATAAAAAAGAAGAAAAAGAAGAATTTCTT AAA	TTAAA	167933		x	
380R	167936	169522	169014	gtgtcagaaatgcaaaggt	AATATTTATATTTTTTATGGAATGTTCCATAA	CATAA	169554	x		
395R	175856	176311	175856	taaaaattcatattaaattgta	ND	.....				
396L	179694	176521	177005	tctccattatctatgtctga	TAATGTTTTAAATGTGGGTGTAGATAGTGGAAAACATTGTATT TTTAAAA	TAAAA	176470		x	
401R	179880	180611	180186	tatgatgtaatggctgtaga	TTTTTAATAGAGCATGCTCTATTA	TATTA	180635			x
415R	184256	184984	184583	gcgtatattcgtgaaggtt	TTTAAATTTTAATGCCATAATTAGCATTA	CATTA	185013	x		
422L	187279	186683	187071	ggaaacattgtgatgaagc	TTTTAATGCTATCTTAGCATTA	CATTA	186661		x	
439L	196182	194569	195084	caggtaggcaaacaccat	ATTTAAAGAAAGTATAATATATATAAATTAAA	TTAAA	194537	x		
443R	196872	204170	203669	tgggacaaatccaaatacaa	ATTTAATGGTTTTCTGACCATTA	CATTA	204193	x		
453L	205047	204631	205025	taacccattacataattaac	ACAAAAATAAAATATAATAAATT	AAATT	204608	x		
457L	206658	205774	206143	tacctataggtgtttatggt	ATTTTTAATGGTTGTTACCATTA	CATTA	205751		x	

**Data presented about the temporal class of the genes were from İnce *et al.* (2013). IE, Immediate early; DE, delayed early; L, late; ND, not detected.

silico analysis (Table 2). The nucleotide distance between the TAATG and CATTA pentanucleotides varied between 5 and 10 nt, with a median of 9 nt (Table S1, Fig. 1B). Three 3′-flanking regions had double CATTA motifs separated with 8 and 1, or 9 and 2 nt from the TAATG sequence. In these cases, the calculations were based on the longest distance. These data suggest that many IIV6 mRNAs may have a 3′-terminal hairpin structure with UAAUG and CAUUA motifs at the base of the stem. For ORF 142R a CATTA motif preceded by a TAATG is present within 70 nt after the stop, and it could be folded into a hairpin structure as well, but this motif was not experimentally confirmed as the 3′-end (Table 1). This might be due to the presence of two types of 3′-end motifs for this ORF that might compete with each other.

2.4. The occurrence of co-variance

If the AUUAG/CAUUA hairpin structure is an important determinant for mRNA 3′-end formation, and possibly for mRNA stability, then changes in the genomic CATTA motif might be compensated by mutations in the TAATG pentamer to preserve secondary structure (co-variance). In the case of ORF 309L, the experimentally determined CTTTA 3′-end of the cDNA was preceded by the complementary TAAAG, and similarly for ORF 380R the CATAA alternative was partnered by TTATG (Table 1, blue sequences). Also for ORF195L the variant CATTAA was accompanied by a TTAATG motif. These typical examples of co-variance further hint towards the functional importance of a secondary RNA structure involving both the UAAUG and CAUUA motifs.

2.5. The length of the inverted repeats extends beyond the UAAUG/CAUUA motifs

The 3′-flanking regions of the ORFs, for which the transcripts were found to terminate after CAUUA motifs, were folded *in silico* using ‘mfold’ software to search for 2D-structures with optimal base-pairing (Zuker, 2003). This exercise showed that the CAUUA motifs indeed had a preference to base-pair with the upstream UAAUG motifs and that the hairpin structure could often be extended beyond this core sequence towards the tip of the hairpin (Fig. 2A, Table S1). The analysis also indicated that in primary transcripts (pre-mRNA) the predicted base-pairing region might start upstream of the UAAUG motif and extend beyond the CAUUA sequence. Thus, the length of the hairpin stem could be extended towards the tip and the base (Fig. 2A). The length of the total putative stem varied per transcript and sometimes side-loops were predicted, either on one or both sides of the stem. The transcripts that showed co-variance in the UAAUG/CAUUA motives (see above) were also folded *in silico* (Fig. 2B), confirming our hypothesis that this co-variance conserves the stem-loop structure. A few transcripts ending with UA-rich motifs were also analysed with ‘mfold’ software, showing that they were also able to form stem-loop structures in the 3′ flanking regions, with the 3′-end motifs in the stem of the hairpins found at the 3′ends of the transcripts (Fig. 2C). This finding is not surprising considering the fact that complementary AT-rich motifs occur with similar frequencies in 3′ flanking regions as TAATG and CATTA motifs (Table 2). Finally, we folded a transcript with a determined CAUUA end motif that was not accompanied by an upstream UAAUG. Surprisingly, in this case, it could form a hairpin with a UAAUG motif downstream of CAAUA, apparently again conserving the recognition as a pre-mRNA cleavage or termination point. This transcript also had an upstream CAUUA motif in its 3′ UTR that appeared to be present in a non-paired loop and was not recognized as the 3′-end (Table 1, Fig. 2D).

2.6. Transcription termination in other iridoviruses

The conservation of RNA processing enzymes prompted us to investigate whether or not the TAATG/CATTA-containing motifs and other structures with the ability to form stem-loop structures were conserved in the 3′- flanking regions of other iridoviruses. We first analysed the pill bug iridovirus (IIV31) isolated from *Armadillidium vulgare* (Order *Isopoda*) (Piégu et al., 2014), which has not yet been assigned to a genus but appears to be the closest known relative to IIV6 (Piégu et al., 2015). The 3′-flanking regions in the IIV31 genome were indeed enriched in both TAATG and CATTA motifs (Supplementary Table S2), appearing 107 and 110 times in these areas with a relative enrichment of 2.9 and 3.0, respectively. The total number of analysed IIV31 ORFs was 203. Visual inspection showed that these two motifs

Table 2
Relative enrichment of penta- and hexanucleotide motifs in the 100 and 200 nucleotides downstream of IIV6 ORFs.

5-mers 100 nt Motif	Occurrence in genome	Real versus expected (%)	Occurrence downstream	Real versus expected (%)	Relative enrichment	5-mers 200 nt Motif	Occurrence in genome	Real versus expected (%)	Occurrence downstream	Real versus expected (%)	Relative enrichment
CATTA	1145	112.5	136	269.2	2.4	**CATTA**	1145	112.5	200	198.0	2.8
ATTAA	2468	101.7	288	239.1	2.4	ATTAA	2468	101.7	449	186.4	1.8
TTAAT	2468	99.0	285	230.2	2.3	TTAAT	2468	99.0	427	172.4	1.7
TAATG	1145	119.1	131	276.2	2.3	**TAATG**	1145	119.9	192	202.4	1.7
TTAAA	3525	145.3	380	315.5	2.2	TTAAA	3525	145.3	586	243.2	1.7
TTTAA	3525	141.4	377	304.5	2.2	TTTAA	3525	141.4	581	234.6	1.7
TAAAA	3331	141.1	327	279.0	2.0	TAAAA	3331	141.1	521	222.3	1.6
TTTTA	3331	130.0	313	245.9	1.9	AAATT	3333	137.4	498	206.7	1.5
AAATT	3333	137.4	307	254.9	1.9	AATTT	3333	133.7	498	201.1	1.5
AATTT	3333	133.7	307	248.0	1.9	ATTTA	2413	96.8	352	142.1	1.5

6-mers 100 nt Motif	Occurrence in genome	Real versus expected (%)	Occurrence downstream	Real versus expected (%)	Relative enrichment	6-mers 200 nt Motif	Occurrence in genome	Real versus expected (%)	Occurrence downstream	Real versus expected (%)	Relative enrichment
TTAATG	527	172.8	111	646.8	4.2	**TTAATG**	527	172.8	141	410.8	2.7
CATTAA	527	179.0	102	573.3	3.9	**CATTAA**	527	179.0	135	379.6	2.6
TAATGG	325	66.2	55	836.6	3.4	**TAATGG**	325	66.2	325	562.8	2.3
TCATTA	398	184.0	61	333.8	3.1	ATTAAA	1236	427.0	1236	314.8.	2.2
CCATTA	325	75.1	49	657.2	3.0	**CCATTA**	325	75.1	325	456.0	2.1
ATTAAA	1236	427.0	278	419.8	2.9	**TCATTA**	398	184.0	398	218.9	2.0
TTTAAT	1236	451.1	176	392.9	2.9	TAAAAT	1071	427.0	1071	240.6	1.9
TAAAAT	1071	427.0	145	342.0	2.7	TTAAAA	1670	427.0	1670	370.3	1.9
TTAAAA	1670	427.0	218	514.1	2.6	AAAATT	1425	427.0	1452	313.7	1.8
TTTAAA	1670	451.1	211	471.0	2.5	TTTTAA	1670	451.1	1670	332.6	1.8

*Only motifs occurring at least 49 times in downstream sequences are shown. Motifs containing CATTA or TAATG are printed in bold and shaded.

again appeared in pairs, suggesting a similar 3′-end strategy as found for IIV6. In addition, also various AU-rich motifs were enriched in these areas, again with ATTAA and TTAAT as most prominent, as in IIV6.

Next, we performed a similar computational analysis for two other insect infecting iridoviruses, *Wiseana iridescent virus* (IIV9, genus *Iridovirus*) and the more distantly related mosquito iridescent virus (IIV3, genus *Chloriridovirus*) (Delhon et al., 2006; Wong et al., 2011). For IIV9 CATTA motifs were found 90 times, TAATG was found 74 times (2.5. and 2.1 times enriched in the 3′regions of the 193 ORFs analysed.). The ATTAA and TAATT were each present more than 200 times (Supplementary Table S2). For IIV3 CATTA and TAATG motifs occurred less frequently (33 and 34 times within the 127 analysed ORFs; enrichment factor of 3 and 2.9). Instead a variety of AT-rich motifs was observed, which appeared to form complementary pairs when 6-nt motifs were analysed: e.g. AATTAA and TTAATT (present 44 and 42 times); ATTAAA and TAATTT (70 and 69 times), TTTTTA and TAAAAA (92 and 91 times). Considering the apparently similar numbers of complementary motifs this may also hint towards a hairpin-based 3′-end determination for viruses in the genus *Chloriridovirus*.

We also included FV3 (Frog virus 3, genus *Ranavirus, Iridoviridae*) in the analysis since it is known that its transcripts are also non-polyadenylated (Willis and Granoff, 1976). For FV3, 5- mer TGTGT and GTGTG were considerably enriched in the 3′ flanking regions and present 44 and 39 times in these areas with a relative enrichment of 5.5 and 4.5, respectively. Also,6-mer TGTGTG and GTGTGT were considerably enriched in the 3′flanking regions and present 31 and 35 times, in these areas with a relative enrichment of 6.2 and 6.0, respectively (Supplementary Table S3). These TG-rich elements may be longer than 6 nt as they also appear in the 7-mer analysis (TGTGTGT (31 times)) and GTGTGTG (35 times) and may in fact reflect the same motif (data not shown), present in various lengths. CATTA motifs were found only eight times, while TAATG motifs were completely absent in these areas. Possibly, the UG-repeat motifs in FV3 transcripts have a similar function as the U/UG-rich regions in cellular pre-mRNAs, although they are not accompanied by poly(A)-motifs. In conclusion, it seems that hairpin-based mRNA 3′-end formation is a conserved feature of invertebrate iridoviruses, but that this feature is absent in ranaviruses.

3. Discussion

It has been known for several years that IIV6 transcripts cannot be converted into cDNA by using an oligo-dT primer (Nalçacioğlu et al., 2003). Similar results have been reported for FV3 (genus *Ranavirus*) (Willis and Granoff, 1976), indicating that iridovirus mRNAs are non-polyadenylated and it has been reported that the dyad symmetry at the 3′ untranslated region of FV3 transcripts may form a hairpin, which may play a role in transcription termination (Rohozinski and Goorha, 1992).

In our current study, we found that transcripts of a considerable number of IIV6 ORFs terminated after a CAUUA motif. In about half of the IIV6 transcripts a RNA hairpin was predicted containing a stem with highly paired UAAUG/CAUUA motifs (Table S1). Although more variation was seen outside of this 10 nt conserved structure, the length of the base-paired stem is often extended towards the tip of the hairpin, leaving only small numbers of unpaired nucleotides in the transcripts (Table S1). The remaining IIV6 transcripts also appear to have the ability to form stem-loop structures at the end of the 3′ UTRs based on complementary AU-rich elements. For eighteen ORFs we found that the transcripts indeed ended after an AU-rich motif (Fig. 2B). We also

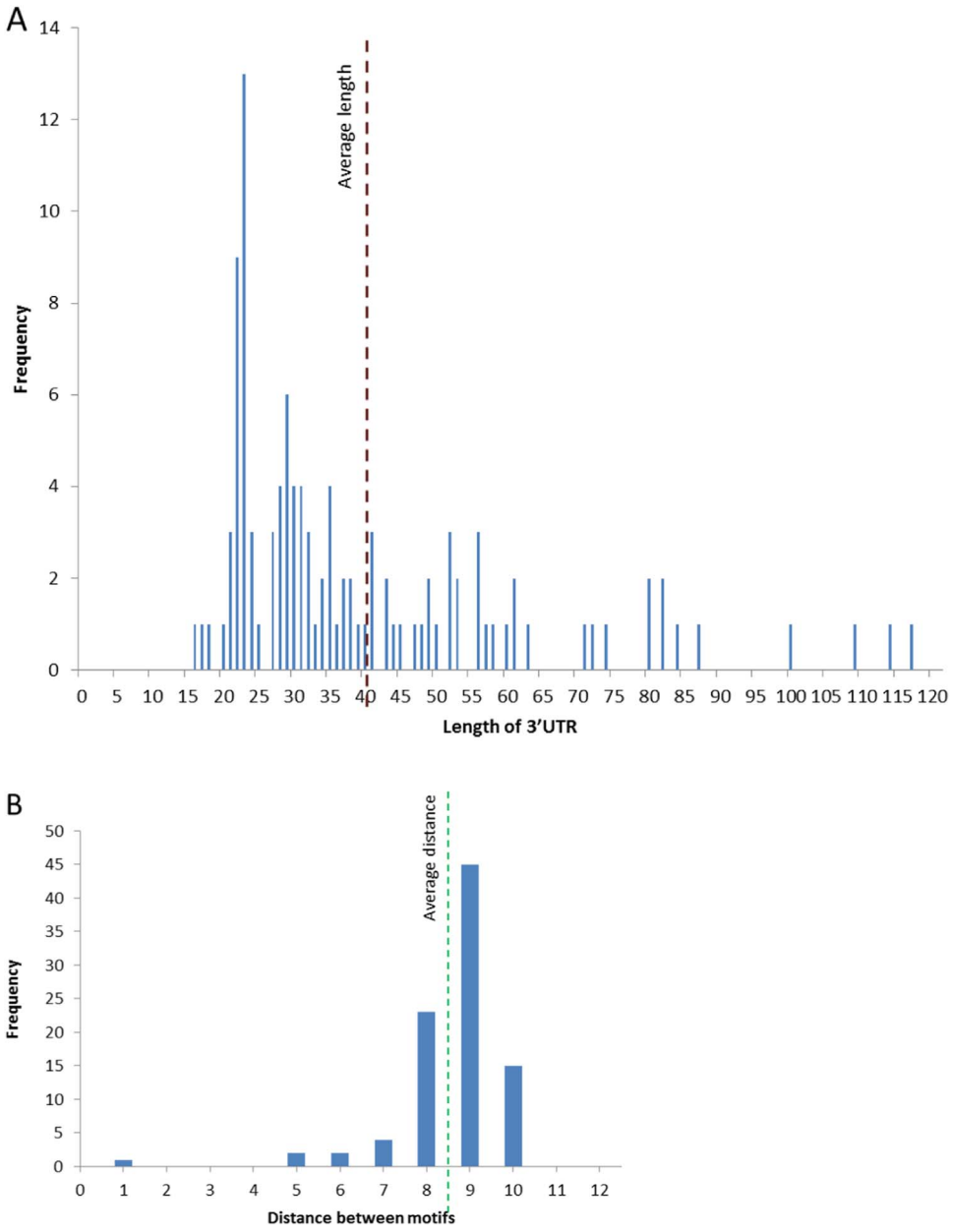

Fig. 1. Variation in the position of the CATTA motif relative to the stop codon and to the TAATG motif. The distribution in length of the predicted 3′ UTR (A) is depicted for the 110 IIV6 transcripts with predicted CATTA end-motifs. The distance between TAATG and CATTA is depicted (B) as the frequency of the appearance of the number of nucleotides (nt) in between these motifs. For ORFs with two CATTA motifs, the longest distance was chosen.

folded the 3′-flanking regions of the primary transcripts of several genes without CATTA motifs in the mfold program and noted that these internally base-paired to a high extent around the mRNA 3′ ends as well (Fig. 2C). When the experimental results for the 3′-ends of the transcripts were compared with the timing of expression of the encoded genes (Table 1), we could not find a correlation between the motif present at the 3′- end and whether a gene was expressed immediate early, delayed early or late in infection. Thus, IIV-6 appears to use a hairpin-based mechanism for mRNA 3′-end formation, where the presence of a CATTA motif, preceded by the complementary TAATG, forms a clear 3′-end signal, that appears to be used for about 50% of the transcripts (see Fig. 3 for a model). Further experiments with constructs carrying mutated 3′ hairpin structures are envisioned to prove the role of these structures in iridovirus mRNA 3′ end formation.

In the viral DNA, the inverted repeats are longer, allowing extension of the base-paired region to the base of the stem in the putative pre-mRNAs (Fig. 2, Fig. 3). Base-pairing appears to be important for 3′-end determination in IIV6 as shown, not only by the striking level of conservation, but also by the existence of co-variance for at least three IIV6 ORFs (CAAUA variants were compensated though nucleotide replacements on the other side of the stem, Table 1, blue motifs). The CAUUA motif as part of the stem-loop structure is likely to determine the pre-mRNA cleavage, or alternatively the direct termination point (Fig. 3). Compared to the histone transcripts, IIV6 shows more variation in the length of the base-paired region as well as in the size and sequence of the loop at the 3′ end of the transcript. There are no unpaired nucleotides following the stem-loop in IIV6 mRNAs, in contrast to the 5 unpaired nucleotides at the 3′ end of histone transcripts. It is therefore questionable whether the CPSF-73 subunit, responsible for histone pre-mRNA cleavage (Dominski et al., 2005; Mandel et al., 2006), also cleaves IIV6 pre-mRNAs. The 3′terminal stem-loop of IIV6 mRNA may stabilise the mature transcripts by protecting them against nuclease attack, comparable to the 3′ end structure of histone mRNAs. It might be that

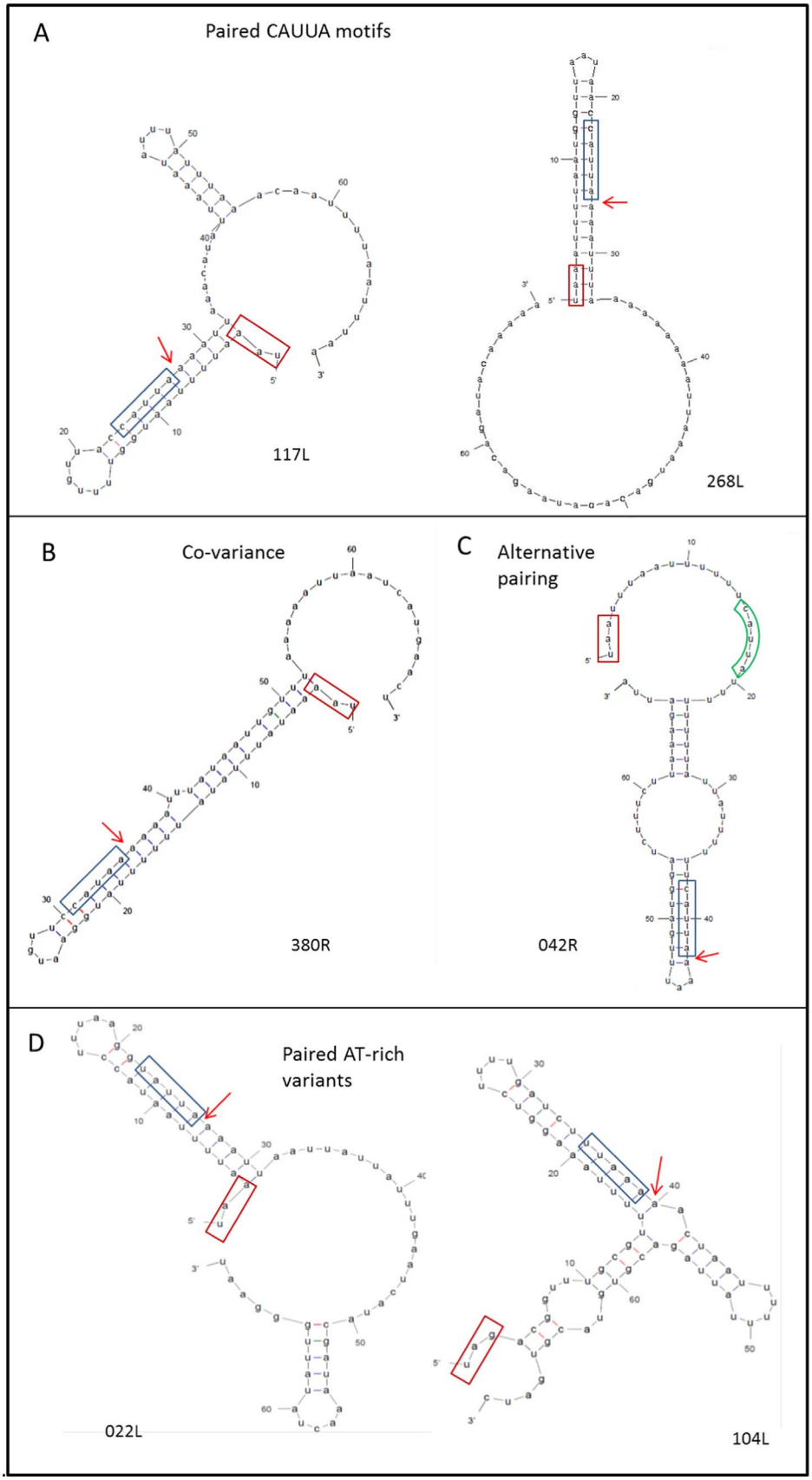

Fig. 2. Examples of predicted base-pairing of putative 3′ pre-mRNA sequences for experimentally analysed transcripts. A 70 nt RNA sequence starting with the stop codon was folded using the mfold webserver (Zuker, 2003). The predicted structures with the lowest energy (ΔG) are shown. A) Base-paired structures for two transcripts (ORF117L and 268L) containing a CAUUA motif (blue boxes) with a complementary upstream UAAUG sequence. B) An example of a transcript with co-variance (380R). D) Two transcripts for which AT-rich 3′ends have been found. C) A 3′ region with CAUUA motif paired with downstream UAAUG (042R). The experimentally determined 3′-ends are indicated with red arrows.

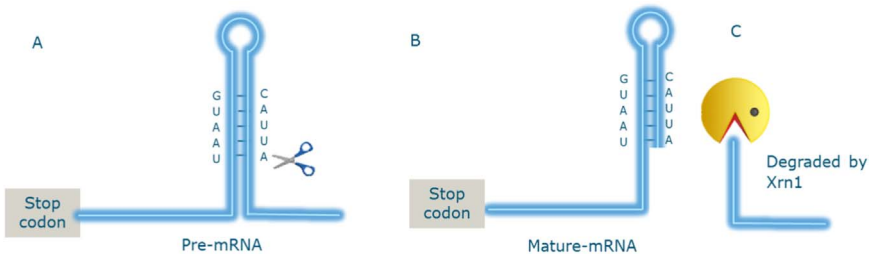

Fig. 3. Model of the 3′ end formation of invertebrate iridovirus mRNAs with UAAUG/CAUUA hairpin structures. A similar model applies for AU-restricted end motifs. A stem-loop structure in the primary transcript (pre-mRNA) might be recognized by a yet unknown enzyme and cleaved after the CAUUA sequence (A) to obtain a mature mRNA (B) that still contains a short hairpin. The remaining 3′ end (C) may be degraded by iridoviral Xrn1 (encoded by IIV6 ORF 012L). In an alternative model (not shown), the RNA polymerase may stop elongating the mRNA after adding the nucleotides of the CAUUA motif in that case the end product is as also as shown in B.

the stem-loop binding protein (SLBP) protects these transcripts from decay, in a similar way as for histone transcripts (Marzluff et al., 2008; Richard and Manley, 2009).

The lack of poly(A)-tails in viral mRNA transcripts is rare, but certainly not unprecedented, particularly in RNA viruses. For example, a select number of positive-strand RNA viruses have 3′terminal RNA structures, e.g. hairpins (Flaviviruses, Turnip crinkle virus) (Markoff et al., 2003; Brinton and Basu, 2015) or tRNA-like structures (*Bromoviridae)* (e.g. Barends et al., 2004). In addition, the mRNAs of the segmented dsRNA viruses, e.g. those in the family *Reoviridae* are also not polyadenylated (McCrae and Woodland, 1981). In general, the lack of polyA tails can be seen as a strategy to allow synthesis of viral proteins under conditions of (partial) shut off of host mRNA translation, but it remains to be investigated whether or not this is also the case in IIV6 infection. The use of alternative 3′-end rules by viruses, rather than those found for most cellular mRNAs may provide the virus with extra strategies to specifically down-regulate host gene expression. Examples are for instance enteroviruses like coxsackie virus (Family *Picornaviridae*) that cleave the poly(A) binding protein PABP (Joachims et al., 1999; Kerekatte et al., 1999), thereby presumably affecting stability and translation efficiency of cellular mRNAs, which is also hampered by cleavage of eIF4G by this picornavirus (Lamphear et al., 1993). eIF4G is needed for translation initiation of capped mRNAs and also an interacting partner of PABP in translation re-initiation. Whether CIV has the ability to interfere with the host mRNA processing machinery or with the proteins interacting with poly(A) tails as host shut off strategies is unclear at this moment, but provides an interesting topic for future investigation.

IIV6 replication starts in the nucleus, while assembly and subsequent maturation of virions occur in the cytoplasm (Goorha, 1982). IIV6 may therefore have access to cellular RNAP II and to cellular RNA processing enzymes, at least in the early phase of infection. Although IIV6 DNA is probably not methylated, in contrast to FV3 DNA (Willis et al., 1989), initiation of IIV6 transcription also requires a (yet unknown) viral activation factor (Cerutti et al., 1989). In later stages, IIV6 transcription and mRNA processing probably rely to a large extent on viral proteins. IIV6 encodes a multi-subunit DNA-directed RNA polymerase (DdRp) for which ORF 176R encodes a homologue of the largest RNAP II subunit (Schnitzler et al., 1994). The DdRp most likely also contains the products of 343L, 428L and 454R (Jakob et al., 2001; Jakob and Darai, 2002). The 107L protein, which also belongs to the RNA polymerase N family (pfam01194), may also be part of this complex as well as a homologue of transcription elongation factor TFIIS (ORF349L). A number of proteins that may assist in processing viral transcripts are also encoded in the IIV6 genome. An example is an Xrn1-exoribonuclease homologue (012L), which may be responsible for degradation of 3′-RNA fragments remaining after cleavage of pre-mRNAs (Fig. 3C), similar to the function of nuclear Xrn-2 (West et al., 2004). The presence of this gene in the iridovirus genome may favour the hypothesis of pre-mRNA cleavage over the idea of direct RNA polymerase termination after the CATTA motif.

The conservation of genes for RNA polymerase and RNA processing enzymes in iridoviruses provided the basis for the search for conserved hairpin structure in 3′flanking regions of in the genomes of other insect iridoviruses (may be with slightly different motifs). First, we inspected the genome of *Armadillidium vulgare* iridescent virus *in silico* and by eye, revealing the frequent appearance of TAATG/CATTA motifs in 3′-flanking regions. The TAATG/CATTA complementary motifs were also observed for about half of the IIV9 3′ flanking regions, and 25% of those of IIV3. Most IIV3 transcripts appear to rely on AU-rich end-motifs that occurred with their complementary counterparts in the genome as shown in the computational analysis. So, hairpin-based 3′ end formation is predicted for invertebrate iridoviruses (genera *Iridovirus* and *Chloriridovirus*). The fact that the enriched 3′-motifs in FV3 (genus *Ranavirus*) have other characteristics (TG rich) may not be very surprising in view of the considerable evolutionary distance between invertebrate and vertebrate iridoviruses (Piégu et al., 2015). FV also does not encode an Xrn1-like RNAse (Tan et al., 2004), in contrast to invertebrate iridescent viruses, further supporting the prediction of a different transcript termination strategy among these two virus groups.

The gene content of iridoviruses varies considerably, suggesting input from various genetic origins. Among the IIV6 genes with homologues in all iridoviruses (the ‘core’ genes) some have CATTA motifs while others have not. In addition, even for homologous genes, the 3′-end motifs may vary between invertebrate iridoviruses. An example is the IIV6 274L gene, encoding the major capsid protein uses the CATTA sequence. In fact, a hairpin structure was already predicted before for the 3′untranslated region of the IIV6 274L transcript (Salem et al., 2008). The IIV9 homologue WIV010R also has a TAATG/CATTA duplet in the 3′-flanking regions, but the corresponding genes in IIV31 (134R) and in IIV3 (MIV014L) lack such motifs.

In conclusion, IIV6 appears to rely at least for about half of its genes on a UAAUG-CAUUA hairpin for mRNA 3′-end formation. The other genes also appear to have transcripts with 3′end hairpins, in this case based on AU-rich paired-motifs. There is no discernible correlation between presence or absence or the type of motif with the timing of expression or whether the gene is an iridovirus core gene or not. The relatively high abundance of paired TAATG-CATTA motifs in the ORF 3′ flanking regions in the genome is not a conserved feature among all invertebrate iridoviruses. These motifs are quite common for the related IIV6 and IIV31 viruses and to a lesser extent for IIV9, but they are not common in IIV3, although some genes carry such motifs also in this virus.

A model for IIV mRNA 3′end formation is presented in Fig. 3. In this hypothetical model, a hairpin will be formed in the primary transcripts and particular sequence motifs in these hairpins determine where the primary transcripts will be cleaved (Fig. 3A). The resulting mRNAs have a shorter hairpin at their 3′ends (Fig. 3B). The rest of the pre-mRNA downstream of the cleavage site may be degraded by nuclease, such as the iridoviral XRN1 (Fig. 3C). Alternatively, the

RNA polymerase may stop as soon as the CAUUA nucleotides have been incorporated in the mRNA. The shorter hairpin that remains in the mRNAs (Fig. 3B) may have a role similar to the poly(A) tail by providing stability to the mature mRNA and/or contributing to translation efficiency. Hairpin formation at the 3′ ends of transcripts appears to be a very common feature for all invertebrate iridoviruses and the fact that the *xrn1* gene is also conserved among IIVs may hint towards a model that involves cleavage of pre-mRNAs.

4. Material and methods

4.1. Viral infection of S2 cells and RNA isolation

The IIV6 isolate was originally from Dr J. Kalmakoff (University of Otago, Dunedin, New Zealand) and kindly supplied by Dr C.J. Funk (USDA-ARS Western Cotton Research Laboratory, USA). Wax moths (*Galleria mellonella*) were used to amplify the virus. The virus titre was determined as described before (İnce et al., 2015). Drosophila Schneider S2 cells were grown in Express Five serum-free medium (Invitrogen) at 27 °C in T25 flasks to a confluency of 80–90% (about 10^7 cells) and infected with IIV6 at a multiplicity of infection MOI > 10 infectious units per cell. Infected cells were harvested at various time points post infection (1, 3, 6, 12, 24 and 48 h). The cells were pelleted at 1500 rpm for 2 min, and washed twice with 0.1 M Tris-HCl, pH 7.6 containing 0.1 M DTT. Total RNA samples were prepared using Trizol reagent (Invitrogen) according to the manufacturer's instructions and subsequently pooled to be able to detect wide range of transcripts from the early to the late transcripts.

4.2. LACE analysis to determine 3′ends of transcripts

To determine the 3′ end of the mRNAs the ligation-based amplification of cDNA ends (LACE) technique was applied as described in detail before (İnce et al., 2013). In short a 5′-phosphorylated DNA oligonucleotide (P1) was ligated to the 3′ end of the RNA molecules to generate a pool of transcripts with a uniform 3′-end. In the next step first strand cDNA synthesis was performed using a primer (P2) that annealed to the uniform 3′-end extensions. In the next step gene-specific primers (See Table 1) were used in combination with P2 that would lead to amplification of approximately 500 bp fragments. After optimization, resulting in unique RT-PCR products per gene, this procedure was performed to analyse the 3′-ends of the transcripts of the 54 IIV6 virion protein genes. The LACE products were sequenced (Macrogen Inc., Korea).

4.3. Determination of sequence motifs enriched in downstream of IIV6 ORFs

In order to identify sequence motifs with a potential regulatory role in mRNA 3′-end formation, the 100 and 200 bp regions downstream of the 211 non-overlapping IIV6 ORFs (Eaton et al., 2007) were analysed for the relative enrichment of sequence motifs, in comparison to the sequence of the genome as a whole using the genome sequence provided by Jakob et al. (2001). This analysis was done in a similar way as before, when screening the sequences upstream of the IIV6 ORFs for conserved promoter elements (Nalçacioğlu et al., 2003). The occurrence of possible 4- to 6-mer sequence motifs in the regions downstream of the ORFs regions was compared to their prevalence in the whole genome taking into account the predicted number of occurrences based on the GC content of the IIV6 genome (Jakob and Darai, 2002) (GenBank: AF303741.1). The analysis was performed using modified versions of the scripts in Perl (http://www.perl.com/) that we described previously (Marks et al., 2006). The scripts were re-programmed in order to be able to calculate with the regions extending from the 3′end of ORFs instead of the 5′ flanking regions. For the homemade data library used in this analysis which originally contained the data for the start and stop nucleotides (nt) of the ORFs, the start was artificially set at the original translational stop codon plus 100 or 200 nt. The stop, which in the original file determines the end of the ORF, was set arbitrary at the stop codon plus 300 or 400 nt, respectively, giving artificial 200 nt sequence blocks that served as artificial ORFs, which were preceded by the 100 or 200 nt present downstream of the real IIV6 ORFs. For the ORFs on the complementary strand, the start was set at the translational stop minus 300 or 400 nt, and the artificial stop at the original ORF stop minus 100 or 200 nt. The artificial ORFs did not interfere with the analysis as no further calculations were based on the sequence of the ORFs (real or artificial). Scripts were run for 4, 5, and 6 nt sequence motifs and only motifs occurring at least 50 times in the 3′-end downstream regions (so at least in approximately 25% of the ORFs) were taken into account. A similar strategy was applied to *Aedes taenriorhunchus iridescent virus* (species *Insect iridescent virus* 3 or IIV3, genus *Chloriridovirus*), *Wiseana iridescent* virus (species *Insect iridescent virus 9* or IIV9; genus *Iridovirus*), an iridovirus isolated from *Armadillidium vulgare* (IIV31; unassigned), and to *frog virus 3* (FV3, a vertebrate infecting iridovirus belonging to the genus *Ranavirus)* (Delhon et al., 2006; Piégu et al., 2014; Tan et al., 2004; Wong et al., 2011). The Genbank numbers for the genomes that we analysed were [NC_008187.1] (IIV3); [NC_015780] (IIV9); [HF92-637.1] (IIV31) and [NC_005946] (FV3).

4.4. 2D-structure predictions

To explore the prevalence of CAUUA motifs over different transcripts and the ability of 3′-flanking regions to form stem-loop structures all the 215 currently assigned IIV6 ORFs (Eaton et al., 2007) were analysed individually by hand for CATTA motifs and upstream complementary TAATG motifs. The location of the CATTA motif was then used to predict the size of the 3′-UTRs. The 3′-flanking regions were also screened for larger inverted repeats that extended beyond the hairpin that is present in the mature transcripts. The program ‘mfold’ (Zuker, 2003) (http://mfold.rna.albany.edu/?q=mfold/rna-folding-form) was used to predict the 2D structure of the 3′-regions of the pre-mRNAs of the IIV6 virion protein genes with default settings.

Acknowledgments

The research was supported by the Scientific and Technological Research Council of Turkey (Grant type 2219). Hendrik Marks is acknowledged for helpful suggestions on how to use the existing promotor motif scripts to find 3′ UTR motifs.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://doi:10.1016/j.virol.2017.06.026.

References

Barends, S., Rudinger-Thirion, J., Florentz, C., Giegé, R., Pleij, C.W., Kraal, B.J., 2004. tRNA-like structure regulates translation of Brome mosaic virus RNA. J. Virol. 78, 4003–4010.
Birnstiel, M.L., Busslinger, M., Strub, K., 1985. Transcription termination and 3′ processing: the end is in site!. Cell 41, 349–359.
Brinton, M.A., Basu, M., 2015. Functions of the 3′ and 5′ genome RNA regions of members of the genus *Flavivirus*. Virus Res., 108–119.
Cerutti, M., Cerutti, P., Devauchelle, G., 1989. Infectivity of vesicles prepared from chilo iridescent virus inner membrane: evidence for recombination between associated DNA fragments. Virus Res. 12, 299–313.
Chen, F., Wilusz, J., 1998. Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. Nucleic Acids Res. 26, 2891–2898.
Delhon, G., Tulman, E.R., Afonso, C.L., Lu, Z., Becnel, J.J., Moser, B.A., Kutish, G.F., Rock, D.L., 2006. Genome of invertebrate iridescent virus type 3 (mosquito iridescent virus). J. Virol. 80, 8439–8449.

Dominski, Z., Yang, X.C., Marzluff, W.F., 2005. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell 123, 37–48.
Eaton, H.E., Metcalf, J., Penny, E., Tcherepanov, V., Upton, C., Brunetti, C.R., 2007. Comparative genomic analysis of the family *Iridoviridae*: re-annotating and defining the core set of iridovirus genes. Virol. J. 4, 11.
Edwalds-Gilbert, G., Veraldi, K., Milcarek, C., 1997. Alternative poly(A) site recognition in complex transcription units: means to an end. Nucl. Acids Res. 25, 2547–2561.
Eliseeva, I.A., Lyabin, D.N., Ovchinnikov, L.P., 2013. Poly(A)-binding proteins: structure, domain organization, and activity regulation. Biochemistry 78, 1377–1391.
Elkon, R., Ugalde, A.P., Agami, R., 2013. Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. 14, 496–506.
Goorha, R., 1982. Frog virus 3 DNA replication occurs in two stages. J. Virol. 43, 519–528.
Gromek, J., Dvir, A., 2010. Eukaryotic gene transcription. In: Sitaramayya, A. (Ed.), Signal Transduction: Pathways, Mechanisms and Diseases. Springer-Verlag, Berlin.
İnce, I.A., Boeren, S.A., van Oers, M.M., Vervoort, J.J., Vlak, J.M., 2010. Proteomic analysis of Chilo iridescent virus. Virology 405, 253–258.
İnce, I.A., Ozcan, K., Vlak, J.M., van Oers, M.M., 2013. Temporal classification and mapping of non-polyadenylated transcripts of an invertebrate iridovirus. J. Gen. Virol. 94, 187–192.
İnce, I.A., Boeren, S., van Oers, M.M., Vlak, J.M., 2015. Temporal proteomic analysis and label-free quantitation of viral proteins of an invertebrate iridovirus. J. Gen. Virol. 96, 196–205.
Jakob, N.J., Darai, G., 2002. Molecular anatomy of Chilo iridescent virus genome and the evolution of viral genes. Virus Genes 25, 299–316.
Jakob, N.J., Muller, K., Bahr, U., Darai, G., 2001. Analysis of the first complete DNA sequence of an invertebrate iridovirus: coding strategy of the genome of *Chilo iridescent virus*. Virology 286, 182–196.
Joachims, M., van Breugel, P.C., Lloyd, R.E., 1999. Cleavage of poly(A)-binding protein by enterovirus proteases concurrent with inhibition of translation in vitro. J. Virol. 73, 718–727.
Kerekatte, V., Keiper, B.D., Badorff, C., Cai, A., Knowlton, K.U., Rhoads, R.E., 1999. Cleavage of poly(A)-binding protein by coxsackievirus 2A protease *in vitro* and *in vivo*: another mechanism for host protein synthesis shutoff? J. Virol. 73, 709–717.
Lamphear, B.J., Yan, R., Yang, F., Waters, D., Liebig, H.D., Klump, H., Kuechler, E., Skern, T., Rhoads, R.E., 1993. Mapping the cleavage site in protein synthesis initiation factor eIF-4 gamma of the 2A proteases from human coxsackievirus and rhinovirus. J. Biol. Chem. 268, 19200–19203.
Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., Tong, L., 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature 444, 953–956.
Markoff, L., 2003. 5′- and 3′-noncoding regions in flavivirus RNA. Adv. Virus Res. 59, 177–228.
Marks, H., Ren, X.Y., Sandbrink, H., van Hulten, M.C.W., Vlak, J.M., 2006. ***In silico*** identification of putative promoter motifs of White spot syndrome virus. BMC Bioinform., 7.
Marzluff, W.F., Wagner, E.J., Duronio, R.J., 2008. **Met**abolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. 9, 843–854.
McCrae, M.A., Woodland, H.R., 1981. Stability of non-polyadenylated viral mRNAs injected into frog oocytes. Eur. J. Biochem. 116, 467–470.
McLauchlan, J., Gaffney, D., Whitton, J.L., Clements, J.B., 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucl. Acids Res. 13, 1347–1368.
Moore, M.J., Proudfoot, N.J., 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136, 688–700.
Nalçacioğlu, R., Marks, H., Vlak, J.M., Demirbağ, Z., Van Oers, M.M., 2003. Promoter analysis of the Chilo iridescent virus DNA polymerase and major capsid protein genes. Virology 317, 321–329.
Piégu, B., Guizard, S., Yeping, T., Cruaud, C., Asgari, S., Bideshi, D.K., Federici, B.A., Bigot, Y., 2014. Genome sequence of a crustacean iridovirus, IIV31, isolated from the pill bug Armadillidium vulgare. J. Gen. Virol. 95, 1585–1590.
Piégu, B., Asgari, S., Bideshi, D., Federici, B.A., Bigot, Y., 2015. Evolutionary relationships of iridoviruses and divergence of ascoviruses from invertebrate iridoviruses in the superfamily megavirales. Mol. Phylogenet. Evol. 84, 44–52.
Richard, P., Manley, J.L., 2009. Transcription termination by nuclear RNA polymerases. Genes Dev. 23, 1247–1269.
Rohozinski, J., Goorha, R., 1992. A frog virus 3 gene codes for a protein containing the motif characteristic of the INT family of integrases. Virology 186, 693–700.
Sachs, A., Wahle, E., 1993. **Pol**y(A) tail metabolism and function in eucaryotes. J. Biol. Chem. 268, 22955–22958.
Salem, T.Z., Turney, C.M., Wang, L., Xue, J., Wan, X.-F., Cheng, X.-W., 2008. Transcriptional analysis of a major capsid protein gene from *Spodoptera exigua ascovirus* 5a. Arch. Virol. 153, 149–162.
Schnitzler, P., Sonntag, K.-C., Muller, M., Janssen, W., Bugert, J., Koonin, E.V., Darai, G., 1994. Insect iridescent virus type 6 encodes a polypeptide related to the largest subunit of eukaryotic RNA polymerase II. J. Gen. Virol. 75, 1557–1567.
Tan, W.G.H., Barkman, T.J., Gregory Chinchar, V., Essani, K., 2004. Comparative genomic analyses of frog virus 3, type species of the genus *Ranavirus* (family *Iridoviridae*). Virology 323, 70–84.
West, S., Gromak, N., Proudfoot, N.J., 2004. Human 5′→3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432, 522–525.
Willis, D.B., Granoff, A., 1976. Macromolecular synthesis in cells infected with frog virus 3. V. The absence of polyadenylic acid in the majority of frog virus 3-specific mRNA species. Virology 73, 543–547.
Willis, D.B., Thompson, J.P., Essani, K., Goorha, R., 1989. Transcription of methylated viral DNA by eukaryotic RNA polymerase II. Cell Biophys. 15, 97–111.
Wilusz, J.E., Spector, D.L., 2010. An unexpected ending: noncanonical 3′ end processing mechanisms. RNA 16, 259–266.
Wilusz, J.E., Freier, S.M., Spector, D.L., 2008. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell 135, 919–932.
Wong, C.K., Young, V.L., Kleffmann, T., Ward, V.K., 2011. Genomic and proteomic analysis of invertebrate iridovirus type 9. J. Virol. 85, 7900–7911.
Yang, L., Duff, M., Graveley, B., Carmichael, G., Chen, L.-L., 2011. Genomewide characterization of non-polyadenylated RNAs. Genome Biol. 12, R16.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucl. Acids Res. 31, 3406–3415.