Transcriptional analysis and genome expression of chicken anaemia virus

Kerry V. Phenix, Brian M. Meehan, Daniel Todd,* and M. Stewart McNulty

Department of Agriculture for Northern Ireland, Veterinary Sciences Division, Stormont, Belfast BT4 3SD, U.K.

Strand-specific riboprobes representative of either strand of the chicken anaemia virus (CAV) replicative form (RF) DNA indicated that only one strand of the RF was transcribed to produce a major 2.0 kb transcript and that the encapsidated DNA strand was of negative sense. Primer extension analysis located a single transcriptional start site at nucleotide position 360 of the CAV sequence. Amplification, cloning and sequencing of the 3' end of the major transcript revealed the polyadenylation site at nucleotide position 21. Northern blot analysis using a series of genomic probes indicated that the 2.0 kb transcript was devoid of splicing and identified a non-transcribed region of the genome. This non-transcribed region was shown to possess promoter activity, enhancing the expression of the human growth hormone reporter gene in a transient gene expression assay. These observations suggest a simple strategy of genome expression involving a functional polycistronic message.

Chicken anaemia virus (CAV) causes clinical and subclinical disease in chickens and is recognized as an important avian pathogen worldwide (McNulty, 1991). The virus has been characterized as a small icosahedral non-enveloped particle (23 to 25 nm in diameter) containing a 2.3 kb circular, ssDNA genome (Gelderblom et al., 1989; McNulty et al., 1990; Noteborn et al., 1991; Todd et al., 1990). A single structural protein (Mr 50K) has been identified in purified virus particles (Todd et al., 1990). It has been proposed that CAV be assigned to the new virus family, Circoviridae, along with two other small ssDNA animal viruses (Studdert, 1993), porcine circovirus (PCV) and psittacine beak and feather disease virus. A circular, double-stranded replicative form (RF) of CAV DNA has been isolated, cloned, sequenced and shown by transfection to contain all the elements required to produce infectious virus (Noteborn et al., 1991; Meehan et al., 1992). Sequence analysis indicated that the three partially overlapping open reading frames (ORFs) encoding potential proteins of Mr 52K, 24K and 13K, together with appropriately positioned transcription initiation and termination signals, could be identified on one of the two strands constituting the RF. In addition, a putative non-coding region rich in transcription regulatory sequences has been located upstream of the three major ORFs (Claessens et al., 1991; Meehan et al., 1992; Noteborn et al., 1991).

In this paper, we report the characterization and mapping of the CAV-specific RNA species and demonstrate promoter activity associated with the non-transcribed region of the genome.

The Cux-1 isolate of CAV (von Bülow et al., 1983), was grown in the Marek’s disease virus-transformed chicken lymphoblastoid MDCC-MSB1 cell line as described previously (McNulty et al., 1988). Total RNA was isolated from infected cells harvested at 12, 24, 36, 48 and 72 h post-infection (p.i.) by the guanidinium isothiocyanate/CsCl method and analysed on a 1% denaturing formaldehyde gel (Sambrook et al., 1989). Virus-specific RNA was identified as early as 12 h p.i. and was shown to increase to a maximum level at 48 h p.i. (data not shown). Subsequent purification of poly(A) mRNA was performed by affinity chromatography using oligo(dT)-cellulose (Pharmacia). A major 2.0 kb transcript together with a minor transcript of approximately 4.0 kb was detected using a random-primed probe prepared from cloned CAV DNA (Todd et al., 1991) and sized using RNA markers (Promega; Fig. 1a). These transcripts were absent from extracts of uninfected MDCC-MSB1 cells and at no time during infection was there any evidence of smaller transcripts (data not shown). Resistance to DNase treatment (data not shown) and its presence in poly(A)-selected preparations support the view that the 40 kb RNA is a virus-specific transcript which may occur if a small proportion of the RNA polymerase molecules fail to terminate transcription efficiently on the circular DNA genome. It is likely that termination of transcription occurs within the region of extreme RNA secondary structure located downstream of the termination codon of the Mr 52K ORF (Claessens et al., 1991).

Strand-specific riboprobes were used to determine the polarity of the encapsidated DNA. The complete 2.3 kb
Fig. 1. Polarity of encapsidated DNA strand and characterization of virus-specific RNA transcripts. (a) Northern blot analysis; total and polyadenylated RNA extracted from uninfected and CAV-infected MDCC-MSB1 cells was separated on a denaturing 1% agarose/formaldehyde gel and detected using a cloned CAV DNA probe. Lanes 1 and 2, total RNA from uninfected and infected cells respectively; lanes 3 and 4, polyadenylated RNA from uninfected and infected cells respectively. The positions of RNA Mr standards are indicated on the right. (b) Clones pCAV14 and pCAV15 were linearized with XhoI to produce run-off in vitro transcripts using T7 polymerase. (c) Labelled sense and antisense in vitro transcripts (probes T7-1 and T7-2 respectively) were hybridized to total RNA extracted from uninfected and infected MDCC-MSB1 cells, and viral ssDNA from purified virions: lane 1, probe T7-2 hybridized to MSB1 RNA; lane 2, CAV-specific RNA; lane 3, CAV DNA; lanes 4 and 5 respectively, unlabelled in vitro transcripts T7-1 and T7-2; lanes 6 to 10, as lanes 1 to 5 above, using probe T7-1.

genome was cloned into the EcoRI site of pBluescript II KS +/− (Stratagene) in both orientations with respect to the T7 promoter. Templates were linearized downstream of the insert by digestion with XhoI and treated with proteinase K at 50 μg/ml (Promega). In vitro transcription reactions using T7 polymerase were performed with 2 μg of template in a 50 μl volume of 40 mM-Tris•HCl pH 7·5, 6 mM-MgCl₂, 2 mM-spermidine, 10 mM-NaCl and 2·5 mM of each rNTP, for 90 min at 28 °C, followed by treatment with DNase. Radiolabelled riboprobes were generated by the inclusion of [³²P]rUTP (Amersham) in the transcription reaction. The constructs (pCAV14 and pCAV15) which were used to produce in vitro T7 transcripts are shown in Fig. 1(b).

CAV ssDNA extracted from virus particles purified by sucrose density gradient centrifugation (Todd et al., 1990) and total RNA from CAV-infected cells were immobilized on nitrocellulose and hybridized to the T7-1 and T7-2 probes in the presence of 50% formamide at 70 °C. Subsequent washing of the blots was performed under stringent conditions (2 × 10 min with 0·1% SSC/0·1% SDS at 20 °C, 3 × 20 min with 0·1%
were separated on a 6% polyacrylamide/8 M-urea gel. Ant sequencing from uninfected (lane 1) and CAV-infected (lane 2) MDCC-MSB 1 cells

Primer extension analysis was performed to locate the 5' end of the transcript. Extended cDNA products initiated using the primer 5' TCTCGGCTAAGCGCCGATTCACTG 3' on 100 gg of total RNA were hybridized to their corresponding unlabelled in vitro transcripts blotted onto nitrocellulose filters. Probe T7-1 hybridized only to ssDNA whereas the T7-2 probe hybridized specifically to viral RNA (Fig. 1 c). This confirms that only one of the two strands of the RF is transcribed to produce mRNA and that the encapsidated DNA strand is of negative sense (Noteborn et al., 1991).

A primer extension assay was performed to locate the 5' start site of the major 2.0 kb transcript. An oligonucleotide primer (5' TCTCGGCTAAGCGCCGATTCACTG 3'), which was complementary to a region approximately 100 nucleotides (nt) downstream from the putative cap site (Noteborn et al., 1991), was 5' end-labelled with 32P using T4 polynucleotide kinase, annealed to 50 μg of total RNA and extended with reverse transcriptase (Sambrook et al., 1989). Primer extension products were fractionated on a denaturing 6% polyacrylamide/8 M-urea gel and a nt sequencing ladder of cloned CAV DNA, initiated using the same primer, was included for comparative purposes. Sequencing was performed using the CircumVent Thermal Cycle DNA Sequencing Kit (New England Biolabs). A single product (Fig. 2 a) was located at nt position 360 (A residue) of the CAV sequence. In agreement with Noteborn et al. (1992), this places the 5' end of the major transcript 30 nt downstream of the TATA box at nt positions 329 to 325. and 25 nt upstream from the initiation codon of the first major ORF at nt positions 386 to 388 of the CAV sequence (Meehan et al., 1992).

Sequence analysis of the CAV genome identified a unique poly(A) addition signal (AAUAAAA) downstream of the three long ORFs at nt positions 2293 to 2298 (Meehan et al., 1992). The precise location of the 3' terminus of the transcript was determined by rapid amplification of 3' cDNA ends, as described by Frohman et al. (1988). Briefly, total RNA (2 μg) extracted from infected and uninfected MDCC-MSB1 cells was independently subjected to cDNA synthesis initiated using the same primer, and 25 nt upstream from the initiation codon of the first major ORF at nt positions 386 to 388 of the CAV sequence.

To investigate the nature of the 2.0 kb transcript, Northern blot analysis was performed using a series of DNA probes spanning the entire CAV genome. Fragments of approximately 200 to 300 bp, produced by cleavage of the recombinant plasmid pCAV14 with a range of restriction endonucleases, were isolated from low melting point agarose (Gibco-BRL) and labelled with 32P (Feinberg & Vogelstein, 1983). The probes were then hybridized to Northern blot strips containing total RNA which had been extracted from infected cells (Fig. 3 a). To summarize, probes A, B, C, D and E (Fig. 3 b), which were representative of the putative coding region encompassing the three overlapping ORFs, hybridized to the major transcript (Fig. 3 a). However probe F, which represented a portion of the putative non-coding region (Fig. 3 b), did not hybridize to the RNA (Fig. 3 a). Given the map positions of the 3' and 5' termini, these results

**Fig. 2. Mapping of the 5' and 3' termini of the major transcript.**

(a) Primer extension analysis was performed to locate the 5' end of the transcript. Extended cDNA products initiated using the primer 5' TCTCGGCTAAGCGCCGATTCACTG 3' on 100 gg of total RNA were hybridized to their corresponding unlabelled in vitro transcripts blotted onto nitrocellulose filters. Probe T7-1 hybridized only to ssDNA whereas the T7-2 probe hybridized specifically to viral RNA (Fig. 1 c). This confirms that only one of the two strands of the RF is transcribed to produce mRNA and that the encapsidated DNA strand is of negative sense (Noteborn et al., 1991).

(b) The PCR product was cloned into the pBluescript IIKS +/− vector and sequenced as described in the text. Positions of the poly(A) addition signal (AAUAAAA) and polyadenylation site are indicated in the sequence.
support the view that the major 2.0 kb transcript is an unspliced message (Noteborn et al., 1992).

Although sequence analysis suggests promoter activity is associated with the CAV non-transcribed region, and the region is rich in sequences of putative regulatory significance, no evidence of promoter activity has yet been demonstrated.

To investigate this further the non-coding region corresponding to positions 2190 to 386 on the circular ssDNA genome (Meehan et al., 1992) was amplified by PCR (94 °C for 1 min; 60 °C for 2 min; 72 °C for 3 min) using the primers 315 (5′ GAACAGCACAAGCTTCTGGGGGACGCAGCC 3′) and 316 (5′ CGCCCTTCCCGATCCCACCCCGGACCGCC 3′) containing artificial HindIII and BamHI recognition sequences respectively. Following restriction endonuclease digestion this PCR product was ligated into the HindIII and BamHI sites of the human growth hormone (hGH) transient gene expression vector p0GH (Allegro, hGH transient gene expression assay system; Nichols Institute Diagnostics) as described by Selden et al. (1986).

The tandem array of four 19 bp repeats present in the non-transcribed region of the CAV genome, which was thought to be characteristic of an enhancer (Meehan et al., 1992), was investigated further by the removal of the first pair of 19 bp repeats (positions 157 to 226) by in vitro mutagenesis using the primer 306 (5′ CAACGAGTGAGACAAAGAGGCGTTCCG 3′). In vitro mutagenesis was carried out using the Muta-Gene (Bio-Rad) M13-based system as described by the manufacturer. Clones containing the entire CAV non-coding region lacking the first pair of 19 bp repeats were characterized by DNA sequencing, subjected to PCR using the primers 315 and 316 and subcloned into the p0GH transient gene expression vector. The constructs containing the entire CAV non-coding region (pCAVGH1) and the non-coding region with the first two 19 bp repeats deleted (pCAVGH2) allow the monitoring of hGH expression arising as a result of promoter activity associated with the CAV non-coding region, and the effect of deletions made within this region, following transfection.

Confluent COS-7 monolayers (75 cm²) were cotransfected as described by Sompayrac & Danna (1981) using 20 μg of either pCAVGH1 or pCAVGH2 in addition to 20 μg of a cytomegalovirus-β-galactosidase construct (kindly supplied by Dr H. Brady, MRC, Mill Hill, U.K.), allowing the transfection efficiency to be estimated. At 72 h after transfection the medium was removed, stored at −20 °C, and the cell monolayer was disrupted by freezing and thawing in 2 ml of assay buffer (50 mM-potassium phosphate, 1 mM-magnesium chloride, pH 7.5).

The concentration of hGH in duplicate 100 μl aliquots of the respective media was determined using the Tandem hGH solid phase, two-site immunoradiometric assay (Hybritec) in accordance with the manufacturer’s instructions. The concentration of β-galactosidase in duplicate 100 μl aliquots of the respective disrupted cell lysates was estimated using the chromogenic substrate chlorophenol red β-D-galactopyranoside (Boehringer Mannheim) as described by Simon & Lis (1987). In addition, total
protein estimations were carried out on duplicate 10μl samples of the respective disrupted cell lysates using a Bradford-based assay (Bio-Rad) allowing the hGH levels in the samples to be expressed in ng/ml after standardizing for transfection efficiency.

Promoter activity associated with the entire CAV non-coding region (pCAVGH1) elicited levels of hGH expression corresponding to 6 to 8 ng/ml. However this level of promoter activity was reduced by 40 to 50% following transfection with the construct pCAVGH2, which lacked the first two 19 bp repeat elements present in the CAV non-coding region. These results provide the first demonstration of promoter activity associated with the non-coding region of the CAV genome and, in addition, highlight the importance of the tandem array of four 19 bp repeats, thought to be an enhancer (Meehan et al., 1992), which is a distinctive feature of the CAV non-coding region.

The identification and mapping of a major 2.0 kb CAV RNA transcript and the promoter activity shown to be associated with the non-transcribed region suggest a novel genome organization and expression for this small DNA virus. Immunoblotting experiments performed in our laboratory have indicated that the three putative ORFs are expressed in CAV-infected MDCC-MSB1 cells (D. Todd, unpublished results), which supports the view that the major 2.0 kb unspliced transcript functions as a polycistronic message as suggested by Noteborn et al. (1992). This simple transcriptional arrangement further distinguishes CAV from PCV, for which at least three types of PCV-specific RNA were reported (Mankertz & Buhk, 1990).

References


(Received 1 September 1993: Accepted 19 November 1993)