Nucleotide sequence and transcriptional analysis of the HindIII P region of a temperature-sensitive mutant of Autographa californica nuclear polyhedrosis virus

Eric B. Carstens* and Albert Lu

Department of Microbiology and Immunology, Queen's University, Kingston, Ontario K7L 3N6, Canada

DNA sequence analysis of the HindIII P region of a temperature-sensitive mutant of Autographa californica nuclear polyhedrosis virus confirmed the specific amplification of 1.4 kb of viral DNA from this region of the genome. The sequenced region included an open reading frame, translated in a counterclockwise direction, which would potentially encode a 74K protein. The amplified DNA was contained within this open reading frame, resulting in in-frame amplifications of a domain within the protein. Transcription studies revealed the presence of a ladder of viral RNA species corresponding to a 2.5 kb transcript carrying tandem repeats of about 1.4 kb. This indicated that the duplicated DNA was transcribed in the same orientation as the p10 gene. We predict that transcripts synthesized from the opposite DNA strand also consist of a ladder of related mRNAs which would be translated to produce a family of p74 proteins with multiple internal domains.

We have previously reported the presence of two genomic variants in stocks of ts8, a temperature-sensitive (ts) mutant of Autographa californica nuclear polyhedrosis virus (AcMNPV) defective for viral DNA replication (Erlandson et al., 1984). ts8A carries insertions of reiterated viral DNA sequences which map near 90 map units (m.u.) on the circular AcMNPV genome. These insertions appeared to consist of amplifications of a specific sequence of viral DNA having homology with the EcoRI P and B fragments. The second genotype, ts8B, has a genomic restriction fragment pattern identical to that of wild-type (wt) AcMNPV. Based on the result that both ts8A and ts8B were equally ts, it was suggested that the genomic alterations of ts8A at 90 m.u. were not located in the ts gene. The ts mutation site of both ts8A and ts8B was physically mapped to the 60 to 62 m.u. region of the genome (Gordon & Carstens, 1984).

The EcoRI P region of AcMNPV is known to encode an abundant very late protein, p10; the DNA sequence of this gene has been determined (Kuzio et al., 1984). Another open reading frame (ORF), potentially encoding a 26K protein, has been identified directly upstream from the p10 coding region (Rankin et al., 1986). In addition, an ORF encoding a 74K protein has been identified downstream of and in the opposite orientation to the p10 gene (Kuzio et al., 1989). Four transcripts have been shown to overlap within the EcoRI P fragment. Transcripts of 750 bp and 2500 bp share a common 5' initiation site upstream from the p10 gene, whereas transcripts of 1100 bp and 1500 bp initiate upstream of the p26 gene (Rankin et al., 1986). The size of the transcript encoding the p74 protein has not been determined due in part to its very low copy number in infected cells.

From restriction enzyme mapping and DNA blot hybridization experiments we postulated that the genomic variant ts8A carried duplications of a 1.4 kb portion of viral DNA from the EcoRI P-B junction region, resulting in a direct repeat of these sequences (Erlandson et al., 1984). Variable numbers of this repeat were present in ts8A genomic DNA packaged into virions such that, when ts8A viral DNA was cut with HindIII or XhoI, the presence of these repeat sequences in different virions resulted in the appearance of a ladder of submolar restriction fragments, each being larger by multiples of about 1.4 kb and homologous to sequences within the HindIII P region (Erlandson et al., 1984). We therefore suggested that ts8A virus stocks actually consisted of a population of genotypes. Some virions carried DNA with a normal HindIII P fragment region while others carried varying numbers of amplified segments within the HindIII P fragment region.

Genetic analyses revealed that the genomic variants were relatively stable, although it was possible to find
plaque isolates from ts8A with restriction patterns like ts8B and vice versa. However, no plaques of ts8A were ever found in which any of the submolar HindIII or XhoI fragments were present in equimolar ratios.

To investigate this interesting mutation further, we sequenced the relevant regions of wt and ts8A DNA. The wt and ts8A HindIII P fragments were subcloned into M13mp19 vectors and the ends were sequenced directly. Internal regions were sequenced by subcloning various restriction fragments of both the wt and ts8A HindIII P fragments into the appropriate M13 vectors and these subclones were then sequenced using a universal M13 primer (Fig. 1a). In addition, we had previously cloned the extra ts8A 1-4 kb EcoRI fragment (pAc8-E1) and demonstrated that it carried the junction site between the repeated sequences (Erlandson et al., 1984). Therefore, this fragment and various subfragments were subcloned into M13 vectors and sequenced (Fig. 1c). We had also previously cloned a submolar 3-6 kb HindIII fragment of ts8A virion DNA (pAc8-01). The insert of viral DNA in this plasmid represents the smallest ts8A submolar HindIII fragment, postulated to contain a single tandem duplication of ts8A DNA within the HindIII P region (Erlandson et al., 1984). Therefore, to ensure that our sequence from ts8A HindIII P and the pAc8-E1 clones was consistent with at least one of the larger submolar HindIII fragments of ts8A which apparently contained tandemly repeated DNA, we also prepared various subclones of the viral DNA insert in pAc8-01 in M13. From the various M13 clones, ssDNA was prepared and sequenced using standard procedures (Sanger et al., 1980) (Fig. 1b).

The sequence of the ts8A HindIII P fragment was determined to be identical to the wt DNA sequence in this region (results not shown). However, the ts8A HindIII P fragment is submolar when isolated from purified virions (Erlandson et al., 1984) and would not be expected to contain any of the rearrangements seen in the family of larger submolar ts8A HindIII fragments, such as the 3-6 kb HindIII fragment cloned in pAc8-01. These larger HindIII fragments were shown by restriction enzyme mapping and blot hybridization to contain rearranged DNA from the HindIII P region (Erlandson et al., 1984). These rearrangements also resulted in the generation of an extra 1-4 kb EcoRI fragment when the large submolar HindIII fragments were digested with EcoRI. Therefore, it was important to sequence this extra 1-4 kb EcoRI fragment. By comparing its sequence with the sequence of the HindIII P fragment (Kuzio et al., 1989), it was determined that this fragment carried the DNA sequence from the EcoRI site at base 566 to base 1605, linked to the DNA sequence from position 207 back to the EcoRI site at position 566. The junction site at base 1605 (Fig. 3b) marked the end of a 1398 bp region which was apparently directly repeated in the 3-6 kb ts8A HindIII fragment in pAc08-1. To confirm this result, two subclones from the plasmid pAc08-1 were constructed. The two subclones were a HindII–HindIII fragment from the left end and a XhoI–EcoRI fragment from the right of the HindIII fragment shown in Fig. 1(b). These two clones were sequenced and shown to carry the expected sequence from the region that was duplicated which was identical to that of the wt sequence (results not shown). These data, together with our previous mapping and hybridization data, provided direct evidence that the submolar bands generated by digestion of ts8A DNA with HindIII are produced by the amplification of a 1398 bp sequence of viral DNA overlapping the EcoRI P-B junction to different copy numbers and packaged into different virions. We were then interested in investigating the possible expression of this amplified DNA.

The HindIII P fragment contains a gene downstream from the p10 gene transcribed in the opposite direction,
which in wt DNA potentially encodes a 74K protein, p74 (Kuzio et al., 1989). The stop codon of p74 is located 12 bp from the stop codon of p10. When the sequence of the ts8A duplication was studied, it was found that the tandem repeat was completely in frame with the p74 reading frame (Fig. 3b). Thus, there was the possibility that both tandemly repeated transcripts and proteins are produced in ts8A-infected cells.

Since it has been previously shown that at least five transcripts are derived from this region of the AcMNPV genome, we investigated the expression of RNA from this region of ts8. Cells infected with ts8A or ts8B were harvested at 24 and 48 h post-infection (p.i.), total intracellular RNA was extracted, denatured with glyoxal and fractionated on agarose gels as previously described (Partington et al., 1990). The RNA was blotted onto nitrocellulose filters and probed with radioactively labelled pAc8-01 DNA which contains the ts8A HindIII P region including one tandem duplication (Fig. 2). Cells infected with ts8B expressed RNA of 0.7 and 2.5 kb, similar to wt virus (Rankin et al., 1986). Much more 2.5 kb transcript was detected than 0.7 kb transcript, but both were clearly present at 24 and 48 h p.i. The 0.7 kb and 2.5 kb transcripts initiate upstream and encode the p10 protein. The 2.5 kb transcript overlaps the p10 coding sequence and terminates near the right end of HindIII P (Rankin et al., 1986). In ts8A-infected cells, both of these transcripts were present but reduced in amount compared with similar sized transcripts in ts8B-infected cells. In addition, a ladder of transcripts of proportional increasing size (2.5, 3.9, 5.3, 6.7, 8.1 kb) and decreasing levels of hybridization were seen. Since the termination site for the 2.5 kb transcript has been mapped close to the right end of the HindIII P fragment at 90.5 m.u., the direct DNA repeats in ts8A would be completely contained within these transcripts. The data indicate that ts8A DNA carries multiple copies of direct repeats which are transcribed into a family of transcripts of increasing size, matching that seen in the ts8A DNA blots. The Northern blots also suggested that the expression of the larger transcripts may cause reduced levels of the normal 2.5 kb and 0.7 kb transcripts because they were detected in reduced amounts in the ts8A-infected cells. This reduction was specific to these particular transcripts because normal amounts of transcripts from, for example, the EcoRI D region were detected in ts8A-infected cells (Fig. 2). This experiment detected RNA transcribed from either strand because dsDNA was used as a probe. Primer extension results suggested that another mRNA is transcribed in a leftward direction and could encode the p74 protein (Kuzio et al., 1989). Our Northern blot analyses suggest that if such a transcript is made in ts8A-infected cells, it would be synthesized in very small amounts but would also consist of a family of related transcripts varying in size by 1.4 kb, corresponding to the size of the direct repeats in the DNA. A model of these transcription products, produced from a ts8A genome carrying one tandem duplication, is presented in Fig. 3(a).

As previously noted, no direct evidence was available to suggest that the alteration in DNA patterns in the HindIII P region was responsible for the ts protein in ts8A. Little is currently known about the function of the gene defective in ts8 mapping to 60 to 62 m.u., although sequence comparisons suggest that it may carry ATP-binding and/or DNA helicase activity (A. Lu & E. B. Carstens, unpublished results). To investigate the possibility that the ts protein was involved in the aberrant replication patterns of ts8A, non-ts recombinants of ts8A and ts8B were selected by cotransfecting purified ts8A or ts8B genomic DNA with a plasmid (pEH3) which carries a 2.7 kb EcoRI–HindIII fragment from the 60.1 to 62.0
Fig. 3. Diagrammatic representation of the 88 to 90 m.u. region of a ts8A genome carrying a single duplication. (a) A restriction map derived from DNA sequence data showing that 1398 nucleotides between the vertical arrows I and II are duplicated between II and III. The effect of this duplication on the size of the 0-7 kb p10 transcript and the 2-5 kb rightward transcript (3.9 kb) are indicated by arrows below the restriction map. Further amplification of this specific region of ts8A DNA in other copies of the genome would result in transcripts of 5-3, 6.7 and 8-1 kb as seen in Fig. 2. Leftward transcripts encoding p74, increasing in size by the same 1.4 kb steps, would also be produced (?). The exact size of the wt p74 transcript is not known but is predicted to be about 2-8 kb. The p10 and p74 ORFs are also shown. The shaded region of p74 corresponds to the direct repeat of the amino acid sequence. (b) Nucleotide sequences of the ts8A regions shown above on the restriction map as I, II and III. The predicted amino acid sequence of the appropriate region of the p74 gene is indicated under the DNA sequence. For each region, the nucleotide number at the right end of the sequence corresponds to the distance from the EcoRI site at the left end of EcoRI P. The amino acid numbers correspond to the amino acid sequence of p74. The junction of the amplified region is indicated with an arrowhead and number corresponding to the nucleotide number from the HindIII site at the right end of HindIII P (Kuzio et al., 1989). Short repeat sequences which may have been involved in the original duplication event are underlined. A short sequence which could form a stem-loop structure is double underlined.

Fig. 4. Southern blot of viral DNA. Total intracellular DNA, purified from cells infected with wt AcMNPV at 25 °C (lane 1) or 33 °C (lane 2), ts8A and ts8B at 25 °C (lanes 3 and 4, respectively), or wt recombinants of ts8, wt8A and wt8B, at 33 °C (lanes 5 and 6, respectively), was digested with HindIII, electrophoresed through agarose gels and hybridized with nick-translated pAc8-01 plasmid DNA. A ladder of related DNA fragments was seen only in ts8A and wt8A-infected cells, confirming that the reiterated sequences were not a consequence of the expression of the temperature-sensitive ts8 protein. Two submolar fragments can be seen in the DNA from wt AcMNPV-infected cells at 33 °C but these represent partial digestion fragments sometimes seen from highly transcribed regions of the genome at this temperature.

prepared from these cells at 24 h p.i. Southern blots of these DNAs, restricted with HindIII, were prepared and probed with DNA from the ts8A HindIII P region (pAc8-01). As shown in Fig. 4, the non-ts recombinants (designated wt8A and wt8B) retained their parental genotype with respect to the alterations in the HindIII P region. This confirmed that the ts protein apparently was not the cause of the amplification of this region in ts8A. None of the non-ts plaques from the ts8A marker rescue stock revealed the absence of the HindIII P fragment and its replacement with a molar concentration of a larger fragment carrying more than one copy of the repeated DNA. These results support the hypothesis that there is an essential gene contained within the region of the repeated DNA which is disrupted in some way when one or more direct repeats of these sequence are present. It is possible that the altered genotypes of ts8A are produced only in the presence of helper virus carrying the normal sequence in this region. On the other hand, the small amount of HindIII P fragment produced by ts8A may
simply result from recombination events which delete extra copies of the repeated DNA in some genomes of ts8A. This second hypothesis is supported by our previous data which demonstrated the generation of authentic HindIII fragment P DNA from the plasmid pAc8-01 (which carries two copies of the repeated sequence) when it was transformed into strains of *Escherichia coli* which were recombination-competent (Erlandson *et al.*, 1984).

The only significant ORF found in the repeated region corresponds to the p74 protein. Recent data suggests that p74 is present in infected cells at late times after infection and that this gene is essential for replication of the virus in insects (Kuzio *et al.*, 1989). The amount of the p74 gene transcript was determined to be very low in infected cells but, based on primer extension results, it was expressed at the same time as the p10 transcripts (Kuzio *et al.*, 1989). Since the duplications in the ts8A genome are within the p74 gene region, and we were able to show that these duplications were transcribed into a family of derivatives of the 2-5 kb RNA, we believe that the mRNA encoded by the p74 gene must also consist of a family of mRNAs carrying duplications of a portion of the p74 gene. Since the duplications of DNA create in-frame repeats with respect to the p74 ORF (Fig. 3b), this strongly suggests that the translation products of these mRNAs in ts8A would consist of a family of proteins carrying more than one copy of an internal domain of the parent protein. However, we do not yet have direct evidence to support this conclusion.

It is still unknown how these direct repeats are generated because the primary sequence of the ts8A DNA in the authentic HindIII P fragment is identical to wt DNA. Several mechanisms have been proposed to account for gene amplification in prokaryotes and eukaryotes (Stark & Wahl, 1984), the most popular of which is the saltatory replication model (Bullock & Botchan, 1981). In this model multiple rounds of initiation near the origin of replication introduces regions of DNA amplification. This process has already been described to account for specific regions of amplified DNA in bacteriophage T4 (Kozinski *et al.*, 1980). This mechanism of amplification, however, requires the presence of a replication origin within or near the amplified DNA. To date no origin of replication within the p74 gene region has been described for AcMNPV.

The most likely mechanism of generation of the multiple copy tandem repeats found in the HindIII P region is an initial duplication event perhaps initiated by the presence of short direct repeats of 16 bp flanking the 1-4 kb DNA region (Fig. 3b). Recombination between short repeats can lead to tandem duplications in *E. coli* by a recA-dependent mechanism (Edlund & Normark, 1981). In bacteriophage, gene duplication has generally been regarded to be the result of illegitimate recombination (Wu & Black, 1987). Recently, a gene amplification event was described for vaccinia virus mutants involving the ribonucleotide reductase gene (Slabaugh *et al.*, 1988, 1989). In this case, a similar sub-stochiometric ladder of DNA fragments was observed corresponding to different restriction fragments possessing varying copy numbers of the basic repeat unit. Generation of these tandem repeats was suggested to be initiated by a non-homologous recombination event leading to duplication of the gene. By a similar mechanism, an initial duplication event in the HindIII P region of ts8A may have been responsible for the generation of arrays of repeat units. Under such a model, homologous recombination between 'out of register' 1-4 kb repeat units would lead to a 'gain and loss' exchange between two DNA molecules carrying tandem duplications. This would result in the generation of sequences of both higher copy number and single copy number as is seen with ts8A genomic DNA (Erlandson *et al.*, 1984).

Sequences surrounding the 1-4 kb repeat unit show a 16 bp imperfect direct repeat. Other structures which may be associated with the flanking regions are imperfect palindromic sequences capable of stem–loop formation (Fig. 3b). Interestingly, this putative structure contains a consensus topoisomerase I recognition sequence -GCTT- (Bullock *et al.*, 1984, 1985) located at the top of the potential stem–loop structure. Non-homologous recombination involving topoisomerase I cleavage has been implicated in simian virus 40 illegitimate recombination (Hasson *et al.*, 1984) and plays an important role in eukaryotic gene amplification (Hyrien *et al.*, 1987). The significance of this structure flanking the 1-4 kb repeat is unknown; however, sequence features such as short homologies, potential stem–loop structures and Alu-like repeats have been noted flanking amplification joints where non-homologous recombination has thought to have played a role in gene amplification (Hyrien *et al.*, 1987).

From these sequence features we suggest that an initial duplication event involving non-homologous recombination was facilitated by the small direct repeats flanking the DNA to be duplicated and occurred in ts8 at an early stage after the initial selection of the mutant (Brown *et al.*, 1979). It is possible that nicks in the DNA caused by topoisomerase I activity resulted in the misalignment of the repeats and were a prelude to the duplication event. Breaks in DNA have been suggested to be generated at greater frequency under conditions where fork movement is severely slowed (Slabaugh *et al.*, 1988). It is not known whether the mechanism of generation of the tandem repeats in ts8A was related to the ts mutation in a gene with putative helicase activities in ts8. However, it
is likely that subsequently, multiple copies of this region were propagated by homologous recombination events. We predict that the ts8A genomes which carry amplified sequences are transcribed to produce a family of mRNAs which would encode a family of related proteins carrying multiple copies of a particular domain of the p74 gene product. We are intrigued by the potential phenotypic expression of such multimeric proteins and are currently investigating these possibilities.

A.L. was supported in part by an Ontario Graduate Scholarship. This work was supported by a grant from the Medical Research Council of Canada.

References


(Received 6 January 1990; Accepted 15 August 1990)