A mutational analysis of the DNA-binding domain of the herpes simplex virus type 1 UL9 protein

Margaret I. Arbuckle and Nigel D. Stow*

MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, U.K.

The herpes simplex virus type 1 origin-binding protein is encoded by gene UL9. We previously described a plasmid, pB1, which encodes a fusion protein containing only the C-terminal 317 amino acids of the UL9 polypeptide and showed that this product retains sequence-specific DNA-binding ability. Two series of pB1 mutants have now been constructed and the polypeptides were tested for origin-binding activity. Using C-terminal truncations, we show that the C-terminal 34 amino acids of UL9 are dispensable for binding and that essential residues lie between positions 801 and 818. Analysis of a series of mutants containing insertions of four amino acids at various positions identified regions of the DNA-binding domain in which alterations either abolished or had relatively little effect upon binding activity. Two mutants which were intermediate in their binding activities also exhibited temperature- or sequence-specific effects.

Introduction

Herpes simplex virus type 1 (HSV-1) encodes seven proteins that perform direct and essential roles in the replication of its 152 kbp DNA genome. One of these, the UL9 protein, binds to specific sequence elements within the viral origins of replication, an event which probably represents the first step towards the initiation of DNA synthesis (for a review of HSV-1 DNA replication see Challberg, 1991).

The HSV-1 origins of DNA replication are specified by two distinct but related sequences. One, oriL, lies close to the centre of the UL region of the genome, and identical copies of the other, oriS, are present within the TRs and IRs inverted repetitions (Stow & McMonagle, 1983; Weller et al., 1985). A distinctive feature of both origins is the presence of a palindromic DNA sequence with a central A+T-rich region. The UL9 recognition sequence, which comprises or is included within YGYTCGCACT (where Y represents C or T) (Koff & Tegtmeyer, 1988; Deb & Deb, 1989; Elias et al., 1990), occurs twice within the regions of dyad symmetry in both oriS and oriL. Within oriS (Fig. 1a) binding site I differs in DNA sequence from site II at only two positions, but has an approximately 10-fold higher affinity for the UL9 protein (Elias & Lehman, 1988). The presence of a functional site I is essential for origin-dependent DNA replication (Deb & Deb, 1989; Weir & Stow, 1990; Hernandez et al., 1991) whereas a low level of origin activity remains when site II is inactivated (Lockshon & Galloway, 1988; Weir & Stow, 1990). The UL9 binding sites within oriL are identical in DNA sequence to the high affinity site I of oriL.

The UL9 protein, in addition to its origin-binding activity (Olivo et al., 1988), also functions as a DNA helicase on partially double-stranded templates (Bruckner et al., 1991; Fierer & Challberg, 1992; Stow, 1992). Energy-dependent unwinding of the origin by UL9 has not, however, been observed (Fierer & Challberg, 1992). The protein exists as a dimer in solution (Bruckner et al., 1991; Fierer & Challberg, 1992) and the 851 amino acid polypeptide chain appears to be organized into distinct functional domains.

Various properties are associated with the region comprising the N-terminal 534 amino acids. The presence of several essential motifs characteristic of a superfamily of proteins that includes many other helicases implicates this portion of the protein in DNA unwinding (Gorbalenya et al., 1989; Martinez et al., 1992). The N-terminal part of the protein is also necessary for efficient dimerization (Elias et al., 1992) and for the observed cooperativity in binding to oriL (Elias et al., 1990, 1992; Hazuda et al., 1992). Interactions involving this region probably also account for the ability of UL9 to distort and/or form complex nucleoprotein structures at the origin (Koff et al., 1991; Rabkin & Hanlon, 1991; Fierer & Challberg, 1992).

The C-terminal one-third of the UL9 protein represents a discrete domain which contains all the information necessary for sequence-specific DNA-binding to the HSV-1 origins. Weir et al. (1989) demonstrated that a fusion protein expressed in Escherichia coli
consisting of amino acids 535 to 851 of UL9 at its C terminus and 260 amino acids from the \textit{Staphylococcus aureus} Protein A at its N terminus interacted specifically with ori, binding sites I and II. Subsequently Deb & Deb (1991) showed by using in vitro translation products that amino acids 542 to 564 and 834 to 851 from within this region were dispensable for binding. Insertions of four amino acids at three positions within the DNA-binding domain abolished binding activity (Deb & Deb, 1991). In this paper we report the construction and analysis of deletion and insertion mutants within the DNA-binding domain of the Protein A–UL9 fusion product described above. The results indicate that mutations within two separate parts of this domain disrupt DNA-binding activity whereas insertions within a third region, which contains a proposed helix–turn–helix motif (Deb & Deb, 1991), have little effect.

**Methods**

\textit{Expression in E. coli.} The parental plasmid, pB1 (Fig. 1 e) consists of an HSV-1 DNA fragment encoding the C-terminal 317 amino acids of the wild-type (wt) UL9 protein inserted into a vector derived from

\begin{table}
\centering
\caption{Plasmids used in these studies}
\begin{tabular}{ccc}
\hline
Plasmid & Linker* & Alteration† & Position‡ \\
\hline
pB1 & - & - & - \\
pD16 & T & 834 LPTE – LP & 20758 \\
pD27 & T & 823 AD1 – AD & 20791 \\
pD32 & T & 818 SANPNA – SASLD & 20807 \\
pD34 & T & 817 SSANPN – STSLD & 20810 \\
pD50 & T & 800 RFKLR – RFKL & 20859 \\
pD62 & T & 788 LLMA – LL & 20896 \\
pI581 & I & 581 TR – TPEFRR & 21519 \\
pI591 & I & 591 AC – ARIC & 21491 \\
pI630 & I & 630 AL – ARIF & 21375 \\
pI652 & I & 652 AR – AGIF & 21308 \\
pI668 & I & 668 GP – GPNIF & 21259 \\
pI691 & I & 691 GH – GPNG & 21190 \\
pI708 & I & 708 RV – RRGN & 21139 \\
pI719 & I & 719 RG – RRNS & 21106 \\
pI799 & I & 799 AR – APEF & 20865 \\
pI838 & I & 838 AW – ARIR & 20750 \\
\hline
\end{tabular}
\footnote*{Termination (CTAGTCTAGACTAG) and insertion (CCCGGAATTCCGG and CCGGAATTCCGGG) linkers are denoted T, I, and I, I, and I, I, respectively.}
\footnote†{Alteration to UL9 amino acid sequence. The number refers to the position of the adjacent amino acid. * represents the C terminus.}
\footnote‡{1 Linkers are positioned immediately downstream of the indicated nucleotide position within the UL9 open reading frame (nucleotides 23261 to 20809 of the HSV-1 genome; McGeoch \textit{et al.}, 1988).}
\end{table}

pRIT2T (Nilsson \textit{et al.}, 1985; Weir \textit{et al.}, 1989). The UL9 sequences are expressed as the C-terminal portion of a fusion protein that contains sequences from the \textit{S. aureus} Protein A molecule at its N terminus. Expression is driven by the bacteriophage \(\lambda\) promoter, \(\lambda p_{\text{r}}\), although we previously induced expression of the fusion protein by raising the temperature of an \textit{E. coli} host specifying a temperature-sensitive repressor, we subsequently found it more convenient to allow constitutive expression in \textit{E. coli} strain D55 (Hanahan, 1985). This strain was used throughout the experiments described here.

\textit{Construction of C-terminal deletion mutants.} Plasmid pB1 was linearized with SalI and digested with nuclelease Bal 31 using the manufacturer’s recommended conditions (Life Technologies). The digested DNA was dephosphorylated and ligated to phosphorylated 14 bp \textit{XbaI} linkers containing termination codons in all three frames (CTAGTCTAGACTAG). The products were cleaved with \textit{XbaI}, recircularized and used to transform \textit{E. coli} DH5 cells to ampicillin resistance. Plasmid DNA from selected colonies was analysed, and the extent of deletion was determined by restriction enzyme digestion and DNA sequencing. The plasmids used in these studies are described in Table 1 (pDr series, where \(n\) indicates the number of amino acids in UL9 that were deleted).

\textit{Construction of insertion mutants.} Plasmid pB1 was partially digested with the enzymes \(Mvnl\) or \textit{HaeIII} and linear molecules were purified from an agarose gel. Twelve bp \textit{EcoRI} linkers were inserted at the site of linearization as described above and the structures of selected insertion mutants were confirmed by DNA sequencing. The linkers inserted at \(Mvnl\) and \textit{HaeIII} sites had the sequences CCGGAATTCCGG and CCGGAATTCCGGG, respectively. The resulting alterations are described in Table 1 (pIn series, where \(n\) represents the amino acid position within the UL9 polypeptide corresponding to the site of insertion).

\textit{Preparation of extracts.} Extracts were prepared essentially as described by Weir \textit{et al.} (1989). Bacteria were shaken in 100 ml L-broth
containing 50 μg/ml ampicillin at 37 °C until an A_{600} of 0.5 was reached. The cells from 50 ml of the culture were centrifuged and the pellets were washed three times with 10 mM-Tris–HCl pH 7.5, 1 mM-EDTA, then were resuspended in 400 μl extraction buffer (20 mM-HEPES-NaOH pH 7.9, 0.6 M-KCl, 25% glycerol, 1 mM-MgCl₂, 0.2 mM-EDTA, 1 mM-DTT, 0.5 mM-PMSE) and extensively sonicated. Supernatants obtained following centrifugation at 11600 g for 15 min were used as extracts. Protein concentrations were determined using Bio-Rad protein assay dye reagent.

**Gel retardation analysis.** The oligonucleotides containing the sequences of binding sites I and II are shown in Fig. 1b and have been described previously (Weir et al., 1989). Ten μg of protein extract in 6 μl extraction buffer was added to an 18 μl mixture containing 80 mM-HEPES-NaOH pH 7.5, 2 mM-DTT, 4 mM-EDTA, 2 μg sonicated calf thymus DNA and 1 ng ³²P-labelled oligonucleotide. After 20 min at the indicated temperature, 5 μl loading buffer was added and the samples were analysed on 5% non-denaturing polyacrylamide gels (Weir et al., 1989).

**Western blot analysis.** Samples of extract containing 1 μg protein were fractionated by SDS-PAGE through 9% gels and electrophoretically transferred to nitrocellulose sheets (Towbin et al., 1979) using Bio-Rad mini-gel kits as recommended by the manufacturer. After electroblotting, the membranes were incubated for 16 h at 37 °C inTBST (10 mM-Tris–HCl pH 8.0, 150 mM-NaCl, 0.05% Tween 20) containing 3% (w/v) gelatin and then incubated with 10 μl TBST containing 1 μl alkaline phosphatase-conjugated goat anti-rabbit IgG antibody (Promega) for 30 min at room temperature. Membranes were washed three times with TBST and the antibody remaining bound (via its Fc region) to the Protein A component of fusion proteins was detected using Promega colour development solution.

**Results**

**Mutagenesis of plasmid pB1**

In order to characterize further the sequence-specific DNA-binding domain of the UL9 protein, mutations were introduced into plasmid pB1 which encodes a fusion protein consisting of amino acids 535 to 851 of UL9 linked to sequences from *S. aureus* Protein A (Weir et al., 1989). A series of pB1 mutants coding for Protein A–UL9 fusions with deletions at their C termini or insertions of four amino acids within the UL9 portion were generated as described in Methods. The mutations within these plasmids and the alterations to the encoded products are documented in Table 1.

**Analysis of pB1 deletion mutants**

Extracts were prepared from untransformed *E. coli* DH5 cells and from transformants carrying the wt plasmid pB1 or plasmids pD16, pD27, pD32, pD34, pD50 and pD62 which express C-terminally truncated forms of the Protein A–UL9 fusion. The ability to bind to a ³²P-labelled probe containing oriₖ binding site I was examined in gel retardation assays. The results are shown in Fig. 2(a). As shown previously (Weir et al., 1989), the extract from cells containing pB1 forms a retarded complex (lane 2) which is absent from untransformed cells (lane 1) or cells transformed with the vector pRIT2T (data not shown). Retarded complexes were also obtained with extracts from cells harbouring plasmids pD16, pD27, pD32 and pD34 but not pD50 or pD62 (lanes 3 to 8).

To determine whether a pB1-related product was present in the extracts from cells containing plasmids pD50 and pD62 a Western blot was probed directly with an IgG–alkaline phosphatase conjugate. This antibody detects the Protein A–UL9 fusion product by virtue of its interaction with the Protein A moiety. Fusion protein was present in all the extracts from transformed DH5 cells, and a gradual increase in electrophoretic mobility was apparent in the samples which correlated with the increasing extent of deletion (Fig. 2b).

These results demonstrate that 34 amino acids can be removed from the C terminus of the UL9 protein without
imparing its ability to interact with binding site I. Since the product specified by plasmid pD50 is non-functional, residues essential for the interaction must lie between amino acids 801 and 818 of UL9.

**Analysis of pBl insertion mutants**

The products specified by the 10 pBl insertion mutants described in Table I were similarly tested for their ability to interact with the binding site I DNA sequence. The results of binding assays performed at 22 °C are shown in Fig. 3(a) and of Western blot analysis of the extracts in Fig. 3(d). Although each of the extracts from cells transformed with a mutant plasmid contained a fusion protein which exhibited an electrophoretic mobility very similar to that of the pBl product, only the pI691, pI708, pI719 and pI838 products were unaffected in their binding ability (Fig. 3a, lanes 8, 9, 10, 12). The products of pI799 and pI630 showed slightly and more severely impaired binding (lanes 11 and 5, respectively) and no binding was detected with the pI581, pI591, pI652 and pI668 proteins (lanes 3, 4, 6, 7). When similar assays were performed at 37 °C the ability of the pI630 protein to form a complex was almost abolished but the relative binding activities of the other proteins were not affected (Fig. 3b).

The ability of the mutant proteins to interact with ori8 binding site II at 22 °C was also tested (Fig. 3c). Comparison of Fig. 3(a) and (c) reveals that the mutant proteins showed similar behaviour towards the two probe fragments with the exception that the pI630 and pI799 products formed much lower amounts of retarded complex with binding site I than site I.

These results, which are summarized in Fig. 4, indicate that four amino acids can be inserted at four separate sites within the DNA-binding domain without affecting the interaction with the UL9 recognition sequence. However, insertions at four other sites abolished detectable interaction and at two further positions generated products that exhibited reduced binding, particularly to the lower affinity site II sequence.

To exclude the possibility that aggregation of mutant fusion protein molecules might account for the observed alterations in binding behaviour, bacterial extracts were also centrifuged at 100000 g for 30 min in a Beckman TLA 100.2 rotor, and the presence of a pBl-related product in the soluble supernatant fraction was examined by Western blotting. All the proteins that either failed to bind to the UL9 recognition sequence (including the pD50 and pD62 products) or exhibited reduced binding were recovered in the supernatant (data not shown). The altered phenotypes of these mutant proteins are therefore
unlikely to result from non-specific aggregation of molecules but rather from specific structural changes.

Discussion

The presence of DNA sequence motifs identical to that recognized by the HSV-1 UL9 protein either adjacent to, or within, A+T-rich palindromes is a feature shared by known or presumed replication origins of other members of the alphaherpesvirus subfamily (Stow & Davison, 1986; Baumann et al., 1989; Nicolson et al., 1990; Robertson et al., 1991; Klupp et al., 1992). In the case of varicella-zoster virus (VZV) and equine herpesvirus 1 (EHV-1), DNA sequence determination has revealed the presence of UL9 gene homologues (genes 51 and 53, respectively; Davison & Scott, 1986; McGeoch et al., 1988; Telford et al., 1992). Furthermore, it has been demonstrated that the C-terminal portion of the VZV counterpart binds specifically to the conserved UL9 recognition sequence (Stow et al., 1990). When the amino acid sequences of the three origin-binding proteins were aligned it was noted that 86 of the 317 residues within the UL9 DNA-binding domain were conserved in both the VZV and EHV-1 homologues. The positions of these residues are indicated on the amino acid sequence of the UL9 DNA-binding domain shown in Fig. 4, which also summarizes the results of our mutagenesis experiments and data from two other laboratories (Deb & Deb, 1991; Martinez et al., 1992).

In general, the results of the different studies are in very good agreement. The work of Deb & Deb (1991) had previously shown that the N-terminal boundary of the domain was most probably located between residues 565 and 596 (although the presence of essential amino acids between positions 535 and 541 was not excluded). The presence of inactivating mutations within the boundary region (I581 and I591) supports this conclusion. Deb & Deb (1991) previously referred to the sequence spanning amino acids 570 to 591 as a pseudo-leucine zipper (LMRNLNSLMGRTRFIYLAGLE, in which the underlined residues form the proposed heptad repeat). Although this region is clearly important for sequence-specific DNA binding, two observations suggest that it is unlikely to act as a conventional leucine zipper. Firstly, the heptad repeat residues are not well conserved in the aligned VZV and EHV-1 sequences (VFKALACPIEQPRLYNTAILGA and LLVELNSPIVREQFVNVAVLGA, respectively), and secondly, amino acids involved in UL9 dimerization are located outside the DNA-binding domain (Elias et al., 1992).
Our studies place the C-terminal boundary of the DNA-binding domain between residues 802 and 818 (defined by mutants D50 and D34) and indicate that the C-terminal 34 amino acids are dispensable for sequence-specific binding. The sequences close to the C terminus (amino acids 839 to 851) represent one of the more conserved regions within the DNA-binding domain. It is possible that these residues may be involved in another function of UL9 apart from that of origin binding.

Three of the insertion mutants that were unaffected in their interaction with sites I and II (1691, 1708 and 1719) are grouped near the centre of the DNA-binding domain and identify a region of the protein which appears to be relatively tolerant of change. Two of these insertions (1708 and 1719) lie within a helix-turn-helix motif (amino acids 698 to 733) predicted by Deb & Deb (1991) using the method of Chou & Fasman (1978). An independent UL9 mutant with a two amino acid insertion within this region was unimpaired in its replicative function (Martinez et al., 1992). It seems therefore very unlikely that the sequences comprising the predicted helix-turn-helix motif are involved in DNA binding. When statistical procedures based on known helix-turn-helix sequences (Dodd & Egan, 1987; Brennan & Matthews, 1989) were used neither this region nor any other part of the DNA-binding domain scored highly (Weir, 1990).

Sequence-specific binding of the UL9 C-terminal DNA-binding domain does not therefore appear to be associated with a single discrete stretch of amino acids. Rather, residues toward either end, separated by at least 200 amino acids, appear to contribute towards this capacity. It is not known whether the inactivating mutations affect parts of the protein that interact directly with the DNA target or whether they operate indirectly through changes in protein conformation. Since the UL9 protein and its VZV and EHV-1 homologues recognize identical sequences, it might be expected that the most conserved parts of the DNA-binding domain would participate in sequence recognition. Unfortunately, mutations within two particularly well conserved segments (amino acids 746 to 762 and 777 to 790) have not yet been examined. It is nevertheless interesting to note that a mutant, 1799, with an insertion close to one of these regions appears to be more greatly impaired in its binding to site II than to site I. More detailed understanding of sequence-specific recognition by UL9 will hopefully follow from studies involving targeted alteration of single amino acids, chemical cross-linking and ultimately X-ray crystallography.

We are grateful to J. H. Subak-Sharpe and A. J. Davison for their helpful criticism of the manuscript, and thank J. McLauchlan for the synthesis of oligonucleotides. M. I. A. was the recipient of an MRC studentship.

References


(Received 14 December 1992; Accepted 15 February 1993)