DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies virus

Andrew K.Cheung

USDA, Agricultural Research Service, National Animal Disease Center, Ames, IA 50010, USA

Received February 24, 1989; Revised and Accepted May 11, 1989 EMBL accession no. X15120

ABSTRACT

The complete DNA sequence coding for the immediate-early protein (IE180) of pseudorabies virus was determined. The coding region of IE180 is 4380 nucleotides for 1460 amino acid residues. G+C content of the non-coding portion of the IE gene is 70.3% while the G+C content of the coding portion is considerably higher at 80.1%. Correspondingly, codons consisting mainly of Gs and Cs are favored. Clusters of amino acid homologies are observed among IE180 of pseudorabies virus, ICP4 of herpes simplex virus type-1 and IE140 of varicella-zoster virus, and are organized similarly in all three polypeptides. Functions exhibited by IE180 are assigned, tentatively, to structural domains of the molecule by analogy to the HSV-1 ICP4 polypeptide.

INTRODUCTION

Pseudorabies virus (PRV) has only one major immediate-early (IE) gene. Two copies of the IE gene are present in the viral genome with one copy in each of the repeat sequences (1−4). The first PRV gene transcribed during a productive infection is the IE gene, presumably by the cellular polymerase II. This gene codes for the transcriptional activation protein of molecular weight 180 kDa (IE180) (5). It is a phosphorylated protein (6) that accumulates in the nuclei of infected cells (7). Experiments with temperature sensitive mutants (tsG) of the IE gene have demonstrated that active IE180 protein is required for the transcription of other PRV genes (5). IE180 can trans-activate heterologous genes such as the human β-globin and adenovirus genes (8,9); in addition, it can also bind to single-stranded DNA (6). Though multifunctions have been attributed to IE180, the intramolecular structure and function relationships remained to be elucidated.

Previous experiments have shown that the PRV IE mRNA is a non-spliced transcript of approximately 5.6 kb (1,2). The 5′ end was mapped and sequenced. The upstream region was shown to contain DNA sequences similar to the enhancer elements of human and murine cytomegaloviruses (1). The 3′ end, including the polyadenylation signal and the actual poly(A) addition site, has also been reported (2).

In this paper, I report the DNA sequence coding for PRV IE180. The deduced amino acid sequence shows striking similarities with ICP4 of herpes simplex virus type 1 (HSV-1) and IE140 of varicella-zoster virus (VZV). All three proteins (IE180, ICP4 and IE140) are immediate-early polypeptides exhibiting similar functions in their respective viral systems. Specific comparisons were made between IE180 of PRV and ICP4 of HSV-1.

MATERIALS AND METHODS

(a) *Recombinant plasmids*. The BamH1 E, I and J restriction fragments of PRV Indiana-Funkhauser (In-Fh) strain were individually cloned into the BamHI site of the Bluescript

plasmid (Stratagene). Plasmid pIEA was generated by inserting BamH1-I 5′ to a cDNA clone that contains BamH1-J sequences and portion of the poly(A) tail. Thus, pIEA contains the entire PRV IE gene but missing the N-terminal 45 untranslated nucleotides (nt) with the 5′ end of the IE gene closest to the T_3 promoter. Plasmids pA122, pA5e and pA144 are 3′ deletion clones of pIEA that terminate at nt positions 4542, 4239 and 4015 of BamH1-I, respectively (Fig. 1). Plasmid pI21 was generated by inserting BamH1-I into the Bam H1 site of Bluescript with the 5′ end of the IE gene closest to the T_7 promoter.
(b) *DNA sequence analysis*. Nucleotide sequence determination was carried out using the dideoxy-chain termination method (10) on double-stranded plasmids (11). Unidirectional deletions were constructed at either end of the PRV DNA present in pIEA by the exonuclease/S1 nuclease method (12). Universal primers were used to prime the sequentially deleted plasmids since the primer binding site were juxtaposed to different regions of pIEA, and the reverse transcriptase sequencing reagents were purchases from Promega. Due to the high G + C content of the IE gene, sequencing artifacts were expected. However, by sequencing closely spaced deletion clones and performing electrophoresis at high temperature (surface temperature of glass plate of the gels was about 65−68°),majority of the artifacts were resolved. At places where artifacts could not be resolved on a particular strand, the complementary strand gave clean sequence data. The nucleotide sequence presented in this report is derived from at least four independent sequencing reactions, two reactions for each strand. Problem areas were usually sequenced up to six times with slightly different clones. DNA sequence was compiled and analysed using the DNASTAR programs (DNASTAR, Inc).
(c) *In vitro transcription and translation*. Plasmids were digested with restriction enzyme PVU II to excise the PRV specific sequences together with the T_7 and T_3 promoters. The digested DNA (1 μg) were used to generate run-off transcripts from the appropriate promoter and the RNAs were then translated in a rabbit reticulocyte system (Stratagene) in the presence of 800 μCi/ml [^{35}S] L-methionine (New England Nuclear). The *in vitro* translation products were analysed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (13).
(d) *Immunoprecipitation*. [^{35}S] L-methionine labeled polypeptides generated in *in vitro* transcription and translation experiments were immunoprecipitated (14) with a normal rabbit serum and an antiserum that was raised against IE180 isolated from PRV infected cells (a gift from Dr. L. Feldman). Briefly, radiolabeled polypeptides were mixed with Protein A-sepharose beads (Pharmacia) that had been pre-bound with serum (1 h at room temperature) in a buffer containing 150mM NaCl, 5mM EDTA, 50mM Tris-HCl (ph 7.5), 1% NP-40, 1% deoxycholate and 1mM phenylmethylsulphonyl fluoride. Bound immune complexes were washed three times with the same buffer and boiled for three minutes before electrophoresis.

RESULTS

(a) *DNA sequence of PRV IE gene*. Mapping of the 5′ and 3′ ends of the PRV IE180 gene has been reported previously (1,2). The genome structure with an expansion of the internal repeat (IR) indicating the restriction enzyme fragments coding for the IE gene is presented in Figure 1. The first 45 nt of the IE mRNA (In-Fh strain) is encoded by BamH1-E (unpublished data), which is identical to that of a PRV isolate reported by Campbell and Preston (1). Minor differences between the two isolates in the remaining 5′ untranslated region (214 nt) have been reported (2). The 3′ end DNA sequence including

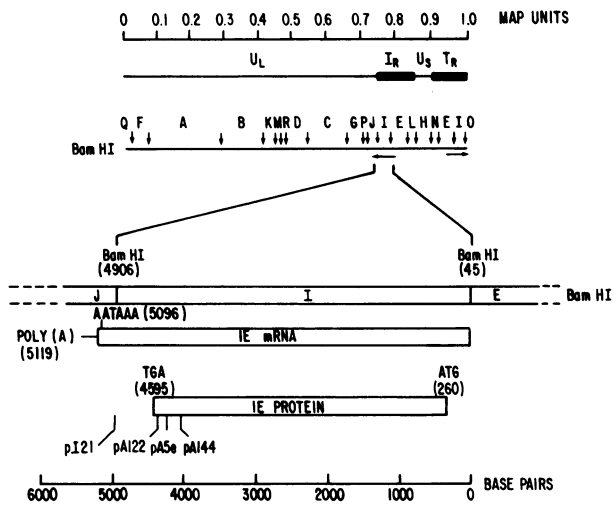

Fig. 1. Organization of the PRV IE180 gene. Locations and directions of transcription of the IE genes are shown at the top. Below the BamH1 restriction enzyme map is an expanded diagram showing the DNA fragments (BamH1-E, -I and -J), the relative position of the mRNA and the IE180 protein. The numbers in parenthesis indicate the nucleotide position measuring from the initiation of transcription. Locations of pI21, pA122, pA5e and pA144 denote the 3′ end of the PRV specific DNA relative to the IE protein in each of the plasmid.

the consensus polyadenylation signal and actual poly(A) addition site was determined from a cDNA clone and it is identical to the genomic DNA sequence present in BamH1-J (214 nt). Since the mRNA is colinear with the viral DNA, BamH1-I must contain the uninterrupted open reading frame (non-spliced) coding for IE180.

In the present work, the entire BamH1-I DNA sequence of 4860 nt was determined, thus the PRV IE mRNA is 5119 nt plus the poly(A) tail. The largest open reading frame with an ATG initiation codon starts at nt position 215 and ends with the termination codon TGA at nt position 4595 (Fig. 2). This open reading frame codes for 1460 amino acid residues. The 3′ untranslated region, including that in BamH1-J is therefore 480 nt, which is considerably longer than its equivalent HSV-1 ICP4 mRNA which has only 65 nt (15). The nucleotide sequence organization for transcription and translation of PRV IE180 is summarized in Figure 1.

(b) *Confirmation of the coding region.* The DNA sequence gave a deduced PRV IE protein of 1460 amino acid residues (Fig. 3) with predicted molecular weight of 153,260 Da, much lower than the SDS-PAGE estimate of 180 kDa. To confirm this observation, run-off transcripts were synthesized from plasmid pIEA and translated *in vitro*. Routinely, seven species of IE180 (mol.wt. 76−180 kDa) were synthesized in addition to an endogenous polypeptide of 48 kDa (Fig. 4a). None of these newly synthesized polypeptides are immunoprecipitable by the normal rabbit serum, but all seven IE180 species are immunoprecipitable with an antiserum specific for PRV IE180. The smaller IE180 species must be derived from prematurely terminated RNAs since they are generated in all the plasmid samples tested.

Due to the high G+C content (80.1%) which may contribute to sequencing artifacts and the lack of A+T rich stop codons to guide the open reading frame, any sequencing

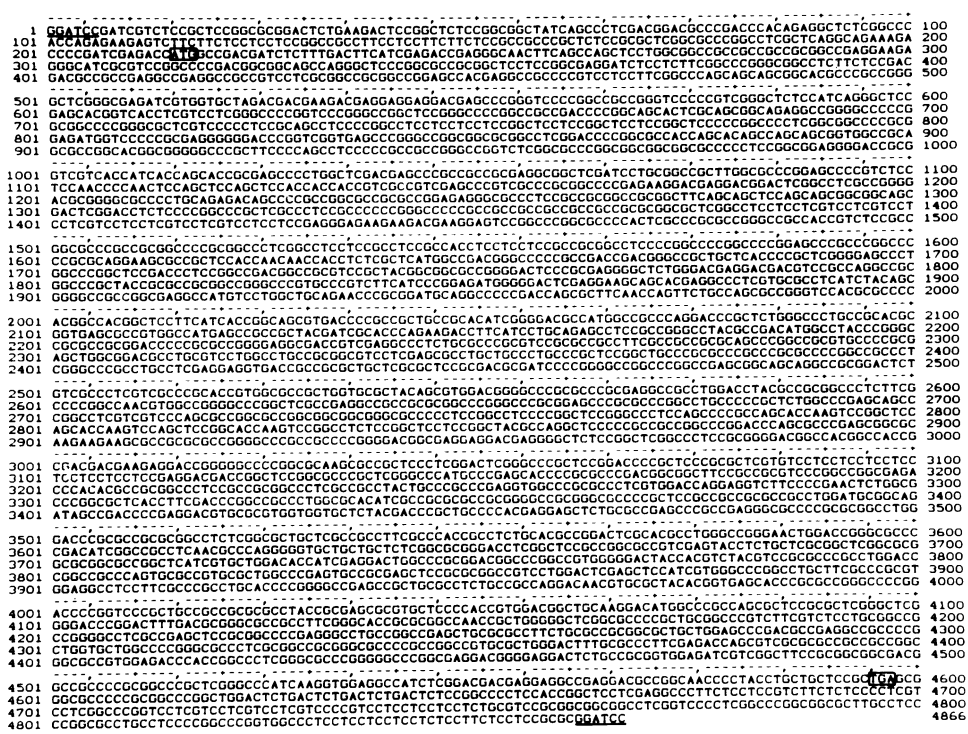

Fig. 2. Complete nucleotide sequence of PRV BamH1-I. BamH1 restriction enzyme recognition sites are underlined at the beginning and end of the DNA sequence. The ATG initiation codon is at position 215 (260 nt from the mRNA initiation site) and the TGA stop codon is at position 4595.

error could easily result in a frame shift that would lead to a different termination point. Specifically, there are two possible stop codons at nt positions 3977 and 4114 present in each of the other two frames. The end point of the coding region was authenticated by *in vitro* translation analysis. Run-off transcripts synthesized from linearized plasmids pIEA, pI21, pA122, pA5e and pA144 were translated in rabbit reticulocyte lysates. Since the DNA sequence data places the termination codon at nt position 4595, plasmid containing a deletion 3′ of the termination codon (pI21) should yield a polypeptide of identical molecular weight (180 kDa) while plasmids with deletions 5′ to the stop codon (pA122, pA5e and

Fig. 3. The deduced amino acid sequence (1460 residues) of PRV IE180.

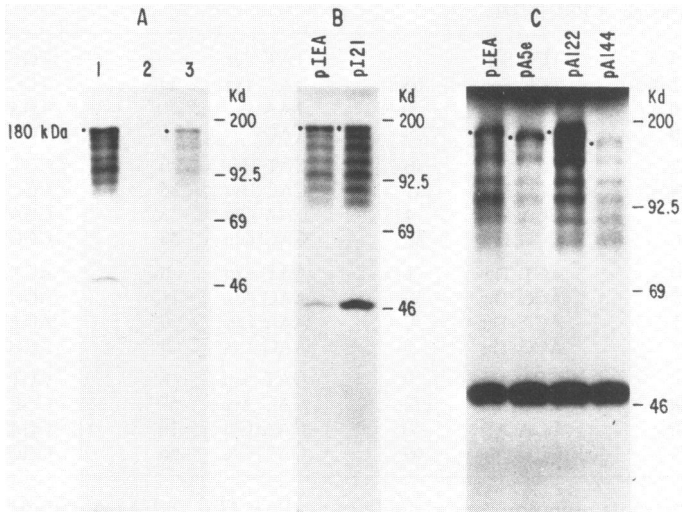

Fig. 4. Analysis of the *in vitro* transcription and translation protein product on 10% SDS-polyacrylamide gel. (A) [^{35}S] L-methionine labeled polypeptides derived from pIEA before immunoprecipitation (lane 1) and after immunoprecipitation with normal (lane 2) and anti-IE180 (lane 3) sera, respectively. (B) [^{35}S] L-methionine labeled polypeptides derived from plasmids pIEA and pI21. (C) [^{35}] L-methionine labeled polypeptides derived from plasmids pIEA, pA5e, pA122 and pA144. (·) denotes the polypeptide with the highest apparent molecular weight and represents the full-length transcription and translation product from each individual plasmid. Molecular weight standards are indicated in kilodaltons on the right of each panel.

pA144) should yield polypeptides of lower molecular weight than the wild-type plasmid (pIEA). Indeed, the full-length *in vitro* translated proteins coded by pIEA and pI21 are identical (Fig. 4b) while the proteins encoded by pA5e and pA144 have lower molecular weight (Fig. 4c). For pA5e, 118 PRV specific amino acid residues were deleted with an addition of 4 plasmid coded residues (a net loss of 114 residues). For pA144, 193 PRV specific residues were deleted with an addition of 4 plasmid coded residues (a net loss of 189 residues). Though pA122 has deletion 5′ of the termination codon, the *in vitro* translation product was indistinguishable from pIEA because only 17 PRV amino acid residues were deleted, and there is an addition of 40 plasmid coded residues (a net gain of 23 residues). These data demonstrated that the stop codon at nt 4595 is the real translation terminator for IE180.

(c) *Base composition and deduced amino acid composition of IE180.* The overall G+C content of the IE180 gene is 78.9%, with the coding region much higher (80.1%) than the non-coding region (70.3%). With such base distributions, the deduced amino acid composition is also skewed to codons rich in Gs and Cs nucleotides (Table 1). The codon usage of IE180 is very similar to that of HSV-1 ICP4 (1298 amino acid residues) encoded by DNA sequences of 81.5% G+C (15).

The deduced amino acid composition is shown in Table 2. Four amino acids each individually exceed 10% of the total amino acid composition, and account for 53.8% of the residues of IE180. Ala is 17.1%, Pro is 13.7%, Arg is 13.0% and Gly is 10.0%. The least common residue are Asn (0.8%), Lys (0.6%) and Met (0.8%). Similar observations have been reported for the corresponding proteins of HSV-1 and VZV. For

Table 1. Codon usage of PRV IE180

TTT Phe	3	TCT Ser	3	TAT Tyr	0	TGT Cys	0
TTC Phe	22	TCC Ser	63	TAC Tyr	17	TGC Cys	18
TTA Leu	0	TCA Ser	0	TAA –	0	TGA –	1
TTG Leu	0	TCG Ser	24	TAG –	0	TGG Trp	17
CTT Leu	2	CCT Pro	16	CAT His	0	CGT Arg	8
CTC Leu	78	CCC Pro	91	CAC His	26	CGC Arg	110
CTA Leu	1	CCA Pro	4	CAA Gln	3	CGA Arg	8
CTG Leu	36	CCG Pro	89	CAG Gln	38	CGG Arg	56
ATT Ile	0	ACT Thr	1	AAT Asn	0	AGT Ser	0
ATC Ile	18	ACC Thr	30	AAC Asn	11	AGC Ser	26
ATA Ile	1	ACA Thr	0	AAA Lys	0	AGA Arg	1
ATG Met	11	ACG Thr	5	AAG Lys	9	AGG Arg	7
GTT Val	0	GCT Ala	9	GAT Asp	6	GGT Gly	3
GTC Val	28	GCC Ala	172	GAC Asp	67	GGC Gly	78
GTA Val	0	GCA Ala	0	GAA Glu	10	GGA Gly	17
GTG Val	34	GCG Ala	69	GAG Glu	66	GGG Gly	48

ICP4, the four most common amino acids (Ala, Pro, Gly and Arg) comprise 54.9% of the total residues (15) and the amino acids below 1% are Asn, Ile, Lys and Met. For VZV IE140, the four most common amino acids are also Ala, Pro, Gly and Arg, and account for 40.7% of the residues (16).

(d) *Comparison of IE180 of PRV and ICP4 of HSV-1.* DNA sequence homology between the genomes of HSV-1 and PRV has been estimated to be approximately 8% by cross-hybridization studies (17,18). One of these homologous regions mapped to the IE genes of HSV-1 ICP4 and PRV IE180 (19–21). Computer search on the National Institute of Health data base (Bethesda, MD) for homologous sequences to PRV IE180 yielded two genes, ICP4 of HSV-1 and IE140 of VZV. All three polypeptides are considered homologs and exhibit common functional properties. ICP4 and VZV IE140 have been shown to share extensive amino acid sequence homology in discrete areas (15). Based on these homologous sequences, the polypeptides have been divided into five intramolecular regions. The deduced amino acid sequence of PRV IE180 was aligned with the ICP4 and IE140 polypeptides. The alignment is to some extent arbitrary. Gaps were introduced whenever necessary to maximize sequence homology. At any event, homologous amino acid sequences are clearly recognized at discrete areas, and five regions of similar structural characteristics can also be established for IE180. Since the comparison between IE180 with either ICP4 or IE140

Table 2. Amino acid composition of PRV IE180

Residue	Number	%	Residue	Number	%
Ala A	250	17.1	Leu L	117	8.0
Arg R	190	13.0	Lys K	9	0.6
Asn N	11	0.8	Met M	11	0.8
Asp D	73	5.0	Phe F	25	1.7
Cys C	18	1.3	Pro P	200	13.7
Gln Q	41	2.8	Ser S	116	7.9
Glu E	76	5.2	Thr T	36	2.5
Gly G	146	10.0	Trp W	17	1.2
His H	26	1.8	Tyr Y	17	1.2
Ile I	19	1.3	Val V	62	4.2

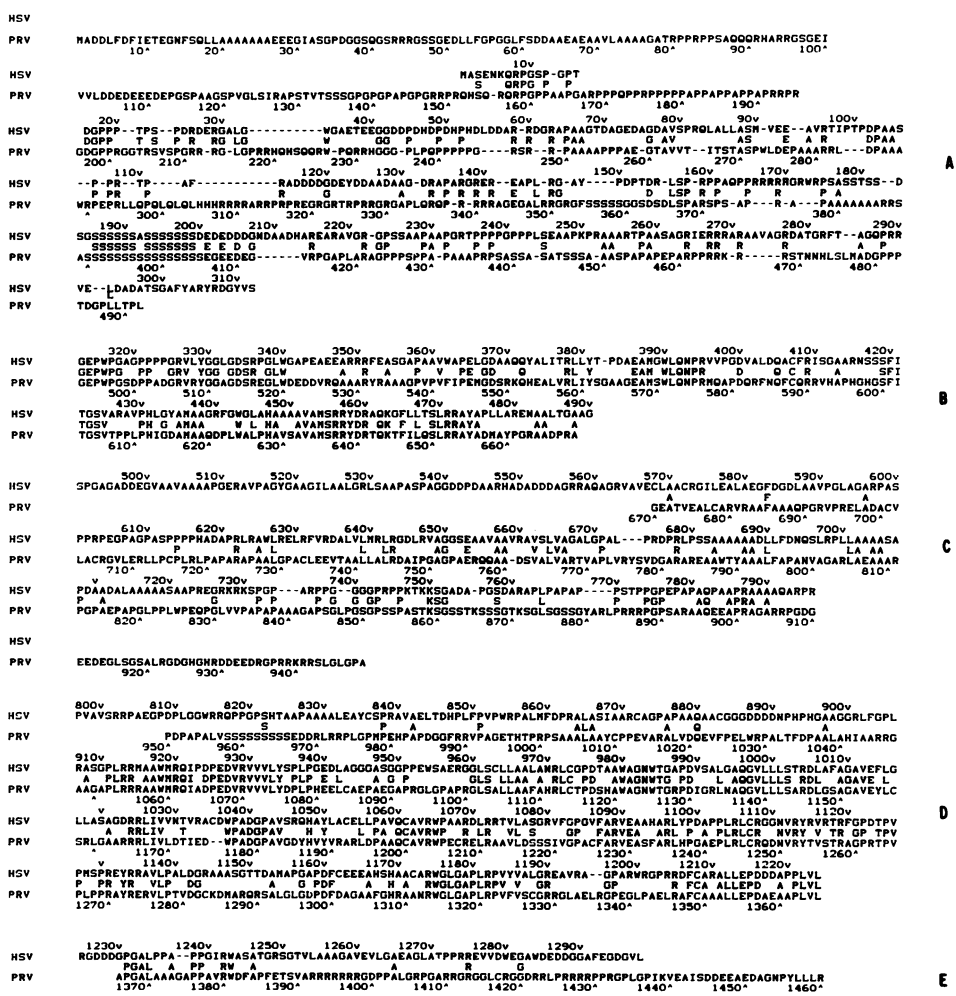

Fig. 5. Alignment of the 5 regions of ICP4 and IE180. Boundary as well as the alignment, to some extent, is arbitrary. Gaps were introduced into the sequence to obtain maximum homology.
(A) Region 1 (ICP4:1−312 vs. IE180:1−492)
(B) Region 2 (ICP4:313−490 vs. IE180:493−699)
(C) Region 3 (ICP4:491−796 vs. IE180:670−949)
(D) Region 4 (ICP:797−1224 vs. IE180:950−1365)
(E) Region 5 (ICP4:1225−1298 vs. IE180: 1366−1460)
Identical amino acids present in the alignment are indicated by the residues between the two sequences.

yielded similar results, only the IE180/ICP4 study is presented (Fig. 5).

Region 1 is a non-homologous region between ICP4 (312 residues) and IE180 (492 residues). The only noticeable similarity is a stretch of serine residues at position 186 of ICP4 and at position 390 of IE180 (Fig. 5a). The size of region 2 is similar for both polypeptides, 177 residues for ICP4 and 176 residues for IE180. It is a highly conserved

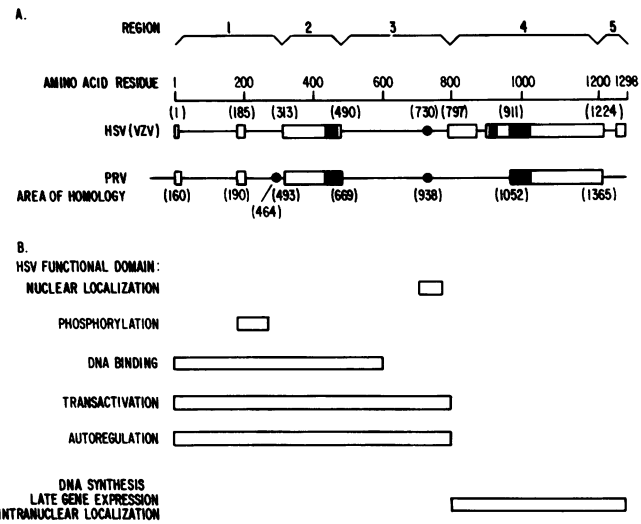

Fig. 6. Comparison of ICP4 and IE180. (A) Summary of the 5 regions established for ICP4 and VZV140 are shown at the top. The PRV diagram is generated by aligning the deduced amino acid sequences of ICP4 and IE180. Areas of homologous sequences are indicated by boxes and those with extensive identical residues are filled in. (·) denotes cluster of positively charged amino acid residues. Numbers in parenthesis indicate the relative amino acid residue position with respect to the polypeptide. (B) The intramolecular structure and function relationship that has been described for ICP4 (22).

area with more than 50% identical residues in the alignment (Fig. 5b). The established size of region 3 is 305 residues for ICP4; however, the IE180 amino acid sequence from 690 to 1051 (381 residues) is non-homologous to region 3 of ICP4. In order to match a cluster of positively charged residues, RKRK at position 727, in region 3 of ICP4, it is arbitrarily decided that IE180 region 3 is 270 residues (from 670 to 949) to include a cluster of similar residues, RRKRR, at position 938 (Fig. 5c). In contrast to a single highly homologous region 4 between ICP4 and VZV 140, region 4 of IE180 can be subdivided into a non-homologous area (between residue 797 to 910 of ICP4 and 941 to 1051 of IE180) and a conserved domain (between residue 911 to 1224 of ICP4 and 1052 to 1365 of IE180) (Fig. 5d). Within the conserved domain, more than 50% of the residues are identical in the alignment. Region 5 is relatively short (68 and 94 residues for ICP4 and IE180, respectively) and it is a non-conserved region (Fig. 5e).

DISCUSSION

Based on the deduced amino acid sequences of ICP4 (1298 residues), IE180 (1460 residues) and IE140 (1311 residues), the predicated molecular weights are 133, 153 and 140 kDa, respectively; instead, the molecular weight estimates on SDS-PAGE are 175, 180, and 185 kDa, respectively. This discrepancy has been attributed to the inability of denaturing gel electrophoresis to size proteins of unusual amino acid composition (15); notably, the high percentage of Ala, Arg, Pro and Gly amino acids.

Striking similarities were observed between the immediate-early genes of PRV, HSV-1 and VZV. Comparison of the deduced amino acid sequence of PRV IE180, HSV-1 ICP4

and VZV IE140 demonstrated clearly that the three polypeptides are closely related. The five-region designation observed between ICP4 and IE140 (15) is also applicable to IE180. Regions 1, 3 and 5 are non-homologous and regions 2 and 4 are highly conserved. The fact that the conserved regions 2 and 4 are separated by the non-homologous region 3, suggests that the conserved domains are critical elements structurally and functionally; and that the distance between these elements may be important. It is interesting to note that short similarities can also be recognized in the largely non-homologous regions. These observations include a stretch of serine residues in region 1 and a cluster of positively charged residues in region 3. In addition, region 1 of VZV IE140 and PRV IE180 are 153 and 180 residues longer than region 1 of ICP4. It appears that a major deletion or insertion occurred to the ancestral gene of these three polypeptides.

Recently, intramolecular structure and function analysis of HSV-1 ICP4 has been carried out and certain properties have been assigned to specific domains of the polypeptide (22,23). Since nothing is known about the correlation of structural domains and functional activities of PRV IE180, comparative studies of ICP4 and IE180 should yield some insight into this relationship. To this end, functional domains ascribed to ICP4 of HSV (22) were incorporated in Fig. 6 for illustration.

Phosphorylation of ICP4 to generate multiple forms of the molecule has been associated with a stretch of serine residues (13 out of 14) at position 185 of region 1. As shown in Fig. 5a, a stretch of 16 serine residues is located at position 390 (region 1) of IE180 which may be involved in phosphorylation.

Functions such as DNA binding, trans-activation and autoregulation have been asssigned to region 2 of ICP4 (22,23). These properties have also been observed with IE180 (5,6,8,9). Due to the extensive homology between this region of ICP4 and IE180 (Fig. 5b), it is likely that the same functions also reside in domain 2 of the PRV polypeptide.

Both ICP4 and IE180 are localized predominantly in the nuclei of infected cells (7,24). Nuclear localization of ICP4 has been attributed to a sequence that consists of a cluster of four strongly positively charged arginine (R) and lysine (K) residues (RKRK) at position 727 of region 3 (15, 22, 23). By analogy, the cluster of RRKRR residues at position 930 of IE180 (region 3) is responsible for nuclear localization; even though, the same cluster is also present at position 464 (region 1) of IE180.

The specific functions associated with regions 4 and 5 of HSV ICP4 have not been fully established. It has been suggested that they may play a role in functions such as DNA synthesis, late gene expression, subnuclear localization and transactivation (22,23). Region 4 of IE180 has a small non-homologous region (Fig. 5d) when compared to ICP4; however, the extensive homology in the rest of the region may be responsible for some, if not all, of the functions mentioned above.

Virtually nothing is known about the intramolecular structure and function relationship of IE180. By analogy to ICP4, several functions can be assigned to various domains of the polypeptide tentatively. Confirmation of these inferences awaits future experimentations. The fact that there is only one PRV IE gene but five HSV IE genes (24) suggests that the IE180 protein may have additional functional domains not yet identified.

ACKNOWLEDGMENTS

I thank L. Patrilla for technical assistance, Drs. P. Kapke and R. Wesley for computer analysis and P. Turner for typing the manuscript.

REFERENCES

1. Campbell, M. E. M. and Preston, C. M. (1987) Virology 157:307–316.
2. Cheung, A. K. (1988) J. Virol. 62:4763–4766.
3. Jean, J. H., Ben-Porat, T. and Kaplan, A. S. (1974) Virology 59:516–523.
4. Rakusanova, T., Ben-Porat, T., Himeno, M. and Kaplan, A. S. (1971) Virology 46:877–889.
5. Ihara, S., Feldman, L., Watanabe, S. and Ben-Porat, T. (1983) Virology 131:437–454.
6. Chlan, C. A., Coulter, C. and Feldman, L. T. (1987) J. Virol. 61:1855–1860.
7. Ben-Porat, T., Kervina, M. and Kaplan, A. (1975) Virology 65:355–362.
8. Feldman, L.T., Imperiale, M. J. and Nervins, J. R. (1982) Proc. Natl. Acad. Sci. USA 79:4952–4956.
9. Green, M. J. R., Treisman, R. and Maniatis, T. (1983) Cell 35:137–148.
10. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74:5463–5467.
11. Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. 152:232–238.
12. Henikoff, S. (1984) Gene 28:351–359.
13. Laemmli, U. K. (1970) Nature (London) 227:680–685.
14. Wesley, R. D. and Woods, R. D. (1986) J. Gen. Virol. 67:1419–1425.
15. McGeoch, D. J., Dolan, A., Donald, S. and Brayer, D. H. K. (1986) Nucleic Acids Res. 14:1727–1744.
16. Davison, A. J. and Scott, J. E. (1985) J. Gen. Virol. 66:207–220.
17. Bronson, D. L., Graham, B. J., Ludwig, H., Benyesh-Melnick, M. and Biswal, N. (1972) Biochem. Biophys. Acta 259:24–34.
18. Ludwig, H. O., Biswal, N. and Benyesh-Melnick, M. (1972) Virology 49:95–101.
19. Ben-Porat, T., Veach, R. A. and Ihara, S. (1983) Virology 127:194–204.
20. Davison, A. J. and Wilkie, N. M. (1983) J. Gen. Virol. 64:1927–1942.
21. Rand, T. H. and Ben-Porat, T. (1980) Intervirology 13:48–53.
22. DeLuca, N. A. and Schaffer, P. A. (1988) J. Virol. 62:732–743.
23. Paterson, T. and Everett, R. D. (1988) Virology 166:186–196.
24. Pereira, L., Wolff, M. H., Fenwick, M. and Roizman, B. (1977) Virology 77:733–749.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.