JOURNAL OF VIROLOGY, Mar. 1984, p. 724–730
0022-538X/84/030724-07$02.00/0

Vol. 49, No. 3

Organization of the Left-Hand End of the Herpes Simplex Virus Type 2 *Bgl*II N Fragment

DENISE A. GALLOWAY* AND MARGARET A. SWAIN

Fred Hutchinson Cancer Research Center, Seattle, Washington 98104

Received 18 July 1983/Accepted 28 October 1983

We have determined the complete nucleotide sequence surrounding the region coding for a 38,000-dalton protein of herpes simplex virus type 2 strain 333 that is encoded between map coordinates 0.58 to 0.595 on the viral genome. The sequence data have revealed an open translational reading frame of 1,011 nucleotides encoding a protein of 337 amino acids. Upstream of those sequences, transcriptional regulatory signals could be identified that overlap another open reading frame of 489 nucleotides, which presumably encodes the carboxy-terminal 163 amino acids of a 140,000-dalton protein. Downstream of the gene encoding the 38,000-dalton protein a polyadenylation signal was found as well as one further along on the other strand near a termination codon that presumably punctuates the gene specifying a 61,000-dalton protein. The DNA sequence data were compared with the amended sequence of the herpes simplex virus type 1 strain KOS 40,000-dalton protein (K. G. Draper, R. F. Frink, and E. K. Wagner, J. Virol. **43**:1123–1128, 1982) and with the 5′ end of the same gene from herpes simplex virus type 1 strain 17 (J. McLaughlan and J. B. Clements, J. Gen. Virol. **64**:997–1006, 1983) to assess the degree of inter- and intratypic variation. The comparison of the sequences has provided a basis for the type specifity of this viral protein and for different mechanisms giving rise to the diversity between herpes simplex virus types 1 and 2.

The discovery that morphological transformation by herpes simplex virus type 2 (HSV-2) can be mediated by sequences contained within the *Bgl*II N fragment (map positions 0.58 through 0.625) (10, 19) stimulated detailed analysis of this region of the genome. By translation of hybrid selected mRNAs four proteins of 140,000 daltons (140K), 61K, 56K, and 35K were shown to be at least partially encoded by the *Bgl*II N fragment (9). More accurate sizing gels have place the molecular weight of the latter at 38K (our unpublished results) and is consistent with the nomenclature of others (5, 26). The 38K protein is the major species late in infection and is also synthesized early before the onset of viral replication. Selection of the message encoding 38K with small fragments from within *Bgl*II-N and with single-stranded DNA segments cloned into M13 vectors placed the mRNA at the left-hand end of the fragment and showed that transcription occurred from left to right.

The polypeptides specified by the homologous region of the HSV-1 genome have also been determined (1, 7) and appear to be arranged identically in both virus types. The 5.0-kilobase (kb) mRNA encoding the 140K protein shares a common 3′ terminus with the 1.2-kb message that encodes 38K, and transcription begins in the adjacent *Bgl*II C fragment around position 0.56 (8). Both messages appear to be unspliced. The region in the HSV-1 genome around the 5′ end of both the 5.0-kb mRNA (7, 18) and the 1.2-kb mRNA (6, 18) have been sequenced; in the case of the 1.2-kb mRNA, the entire DNA sequence has been obtained (6). In this report we characterize the sequences specifying the HSV-2-encoded 38K protein.

One interesting feature of the 38K protein is that it has been reported to be extremely type specific (5). Rabbit anti-HSV-1 serum was not able to precipitate the HSV-2 38K protein from extracts of cells labeled in vivo or from total or selected RNA translated in vitro (5), whereas anti-HSV-2 serum readily precipitated the 38K protein. We have also demonstrated that monoclonal antibodies directed against the HSV-2 38K protein are extremely type specific (9, 14). The two monoclonal antibodies directed against the 38K protein also react, although with various affinities, with the 140K protein.

The function of the 38K protein is not entirely clear. Our studies (11) indicate that the sequences encoding the 38K protein are not needed to initiate morphological transformation, although the presence of the putative 38K protein in HSV-2-transformed cells has been reported (26). In one study (12) women with cervical carcinoma were reported to have significantly elevated levels of antibody to a HSV-2 specific 38K protein, presumed to be the protein encoded by *Bgl*II-N. The 140K protein (also known as ICP6) may also be equivalent to the protein described as ICP-10, which has been detected in cells transformed by the HSV-2 *Bgl*II C fragment (16) as well as on cervical carcinoma cells (2). A more recent study (15) has shown that the viral enzyme ribonucleotide reductase maps to this region of the genome, and that antibody raised against the enzyme precipitates a 140K and a 38K protein, suggesting that one or both of these polypeptides are viral enzymes.

MATERIALS AND METHODS

Construction of deletion fragments. Construction of the recombinant plasmid pP2 derived from pDG401 was described previously (9). The plasmid pmtrIIa was constructed by cleaving P2 with *Pst*I (Bethesda Research Laboratories, Inc.), removing the overhanging nucleotides by digestion with S1 nuclease (Boehringer Mannheim Biochemicals), ligating *Hin*dIII linker molecules (Collaborative Research, Inc.) to the ends, cleaving with *Hin*dIII and *Bam*HI to release the viral insert, and recloning the fragment into a pBR322 vector that had been cleaved with *Hin*dIII and *Bam*HI (New England Biolabs). To create deletions from the *Bam*HI site, 20 μg of pmtrIIa was linearized by digestion with *Bam*HI (Bethesda Research Laboratories), followed by digestion at 21°C with 600 U of exonuclease III (New England Biolabs) in a total volume of 200 μl containing 0.066 M Tris (pH 8.0) and 0.66 mM $MgCl_2$ as described by Sakonju

* Corresponding author.

et al. (20). A volume of 10 μl was removed at 15 s intervals for 0 to 20 min and pooled into 80 μl of 10× S1 buffer (0.5 M sodium acetate [pH 4.5], 1.5 M NaCl, 5 mM $ZnSO_4$). To this pool 580 μl of water was added with 250 U of S1 nuclease and incubated for 15 min at 21°C followed by 10 min at 0°C. After incubation, 100 μl of buffer containing 0.5 M Tris (pH 9.5) and 100 mM EDTA was added to the mixture to stop the reaction. The sample was phenol extracted twice, ether extracted three times, and ethanol precipitated. The pellet was extensively washed with 70% ethanol. A 20 M excess of *Bam*HI linker molecules labeled by polynucleotide kinase with [^{32}P]ATP was ligated to the pool of shortened fragments in a 30-μl reaction containing 60 mM Tris-hydrochloride (pH 7.6), 5 mM $MgCl_2$, 0.33 mM spermidine, 5 mM dithiothreitol, 10 mM ATP, and 4 U of T_4 ligase (Bethesda Research Laboratories) at 4°C for 12 h. After ethanol precipitation the pellet was suspended in 10 mM Tris (pH 8.0) and 1 mM EDTA and digested with 40 U each of *Hin*dIII and *Bam*HI in a buffer containing 60 mM NaCl, 6 mM Tris (pH 7.6), and 6 mM $MgCl_2$ for 4 h at 37°C. After digestion the mixture was sized by electrophoresis on a 1% agarose gel, and DNA fragments ranging in size from 2.0 to 0.5 kb were recovered by electroelution into hydroxyapatite. These fragments were ligated to *Hin*dIII-*Bam*HI–treated pBR322 vector, and the resultant molecules were transformed into *Escherichia coli* HB101. Colonies were isolated that were ampicillin resistant and tetracycline sensitive and screened for size by digestion with restriction enzymes.

Construction of recombinant M13 phage. The viral fragments from the left end of *Bgl*II-N were derived from independently isolated plasmids of HSV-2 DNA; that is, the *Bgl*II-*Bam*HI fragment was derived from a clone containing the *Bam*HI E fragment, the *Bgl*II-*Sst*I fragment was derived from the *Bgl*II N fragment, and the *Bam*HI-*Hin*dIII fragment was derived from pmtrIIa. These fragments were cleaved with a variety of restriction enzymes and ligated into the approximately cleaved replicative form of the vectors M13mp7, mp8, mp9, mp10, and mp11. The details are in the legend to Fig. 1. The DNAs were used to transfect *E. coli* JM101 and plated onto agar plates in the presence of isopropyl-thiogalactoside and 5-bromo-4-chloro-3-indolyl-D-galactoside. The recombinants, which were detected as white plaques, were picked and grown for 6 to 8 h with *E. coli* JM101, and the phage were purified from the culture supernatant by precipitation with 10% polyethylene glycol. The DNA was released from the phage by deproteinization with proteinase K (10 μg/ml) and purified by phenol extraction and ethanol precipitation. The identity of the recombinant phage was determined by restriction analysis and sizing of the replicative forms on agarose gels, hybridization of the phage DNAs to each other to establish pairs, and DNA sequence analysis.

DNA sequencing methods. Sequence analysis by dideoxynucleoside triphosphate chain termination (22) was performed by utilizing as a template the single-stranded M13 phage DNAs (23). A 17-residue fragment (a gift of R. Gelinas) that hybridized adjacent to the cloning site was used to prime the DNA synthesis reaction, which was extended by Klenow polymerase (New England Nuclear Corp.) in the presence of [^{32}P]dCTP and [^{32}P]dTTP, unlabeled dATP and dGTP, one dideoxynucleotide in each of four reactions, for 15 min at room temperature as previously described (27). Thin sequencing gels (21) containing either 6 or 8% polyacrylamide were used to resolve the products. The gels were exposed for 12 h on XAR-5 X-Omat film.

Deletions generated by the method described above were sequenced by the method of Maxam and Gilbert (17). Approximately 10 μg of each deletion fragment used for sequencing was digested with *Bam*HI and labeled at the 3′ end with 10 U of reverse transcriptase (Life Sciences Inc., St. Petersburg, Fla.) in the presence of [^{32}P]dCTP and [^{32}P]dTTP, 0.5 mM dATP and dGTP, 5 mM dithiothreitol, 60 mM NaCl, 6 mM Tris (pH 7.8), and 6 mM $MgCl_2$. The labeled mixture was ethanol precipitated, resolved by electrophoresis on a 1.2% agarose gel, and electroeluted into hydroxyapatite. Chemically degraded DNA fragments were sized on thin gels consisting of either 8 or 18% polyacrylamide in 100 mM Tris-borate [pH 8.3]–2 mM EDTA. The nucleotide sequence data from each template was entered into a VAX/VMS computer (Digital Corp.), and the sequence was compiled and analyzed by using previously established programs (13, 24, 25).

RESULTS

To obtain the complete nucleotide sequence of the segment of DNA encoding the HSV-2 38K protein, we sequenced the region from the *Bgl*II site (map coordinate 0.58) to the *Pst*I site (map coordinate 0.60). Two approaches were used. In the first, a fragment from the *Bam*HI site (map coordinate 0.585) to the *Pst*I site (map coordinate 0.60), designated pmtrIIa, was used to generate a series of deleted fragments constructed in vitro, and each fragment was sequenced directly after chemical degradation (17). A summary of the collection used to obtain the sequence in this study is shown in Fig. 1. In the second approach restriction enzyme fragments were cloned into M13 vectors, and the single-stranded phage DNAs were used as templates for sequencing by the dideoxynucleoside triphosphate chain termination method (22). Because pDG401 was found not to contain the authentic viral sequences from the *Bgl*II site to the *Bam*HI site (Galloway and Swain, unpublished results), this fragment of DNA was obtained from a plasmid containing the *Bam*HI E fragment. The *Bgl*II-*Bam*HI fragment was cloned in its entirety in both orientations in M13mp7, and a library of *Sau*3A fragments was made. The viral sequences from pmtrIIa were cleaved with a variety of restriction enzymes and cloned into appropriate vectors as described in the legend to Fig. 1. To obtain sequence data spanning the *Bam*HI site, the complete *Bgl*II fragment was recloned, and the *Bgl*II-to-*Sst*I fragment spanning the *Bam*HI site was cloned into M13mp10. The collection of M13 recombinant phage used is shown in Fig. 1. The data obtained from the two methods of DNA sequence analysis collectively provided a minimum of one reading from both strands throughout the entire fragment.

The nucleotide sequence of the noncoding strand of the HSV-2 38K gene is shown in Fig. 2 beginning at the *Bgl*II site (map coordinate 0.58) and extending 1,890 nucleotides rightward of that site along the prototype arrangement of the genome. Nucleotides 100 through 1,625 overlap those reported for the HSV-1 strain KOS 38K gene (6), and nucleotides 40 through 751 overlap those reported for the HSV-1 strain 17 5′ end of the 38K gene (18). For heuristic reasons we will divide our results into three sections: the 5′ end, that is, those sequences that fall between the *Bgl*II site and the putative start of translation of the 38K protein (nucleotides 547 through 549); the coding region, from the initiation site (nucleotides 547 through 599) to nucleotide 1,557, which just precedes the TGA termination codon; and the 3′ end, extending from the termination codon (nucleotides 1,558 through 1,560) for another 330 bases.

5′ region. Analysis of the precise start of transcription for

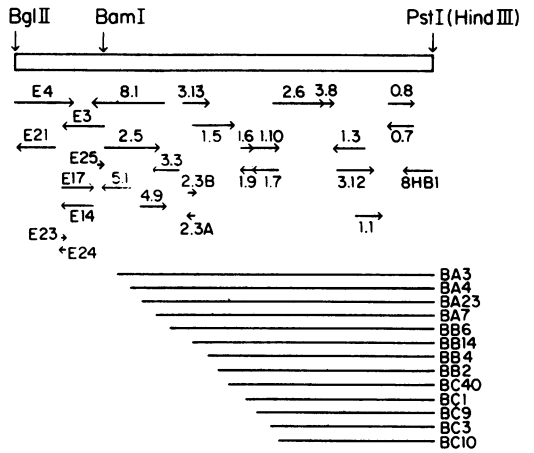

FIG. 1. Strategy used to sequence the left-hand end of the *Bgl*II N fragment. Three fragments of viral DNA were used to cover this region. The viral fragment from pmtrIIa was cleaved with restriction enzymes and inserted into appropriately cleaved M13 vectors as follows: 0.7, 0.8, 1.0 series, and 2.0 series, *Sau*3A fragments into M13mp7-*Bam*HI; 3.0 series, *Hpa*II fragments into M13mp7-*Acc*I; 4.0 series, *Sst*I fragments into M13mp10-*Sst*I; 5.0 series, *Sst*I-*Bam*HI fragments into M13mp11-*Sst*I-*Bam*HI; 8HB1, *Hin*dIII-*Bam*HI entire fragment into mp8 *Hin*dIII-*Bam*HI. The *Bgl*II-to-*Bam*HI fragment was prepared from a plasmid containing the *Bam*HI E fragment and was cloned in its entirety in both orientations in *Bam*HI-cleaved M13mp7 (E3 and E4). In addition the fragment was cleaved with *Sau*3a and inserted into *Bam*HI-cleaved mp11 (E5 through E28). The entire *Bgl*II N fragment was recloned into pKC7, cleaved with *Bgl*II and *Sst*I, and cloned into M13mp10. The phage 8.1 contains the left-hand *Bgl*II-*Sst*I fragment. The direction of the arrow indicates which strand was inserted, and the length of the arrow indicates the amount of DNA sequence that was obtained, not necessarily the size of the fragment. The deletion fragments used for Maxam-Gilbert sequencing, which were made by removing nucleotides from the *Bam*HI site, are shown at the bottom of the figure.

the HSV-2 1.2-kb message (18a) placed the site at nucleotide 410 at exactly, or within a few nucleotides of the start of the HSV-1 1.2-kb message (6, 18). Located upstream (nucleotides 382 through 388) of that was the sequence CATATAA, which could serve as the TATA homology (3) and was identical in HSV-1 and HSV-2. In most eucaryotic messages a CAT homology is found 80 to 90 bases upstream from the start of transcription (4). No recognizable signal was found in the HSV-2 sequence; however, located upstream from the start of transcription of the 1.2-kb RNA, a 17 mer GAGAAGGCGGACGGGAC (nucleotides 327 through 343) was identical in HSV-1 and HSV-2.

The first ATG following the start of transcription was located 140 bases downstream at positions 547 through 549. The transcribed leader sequence is very interesting. Within the first 82 nucleotides, the number of nucleotides was identical to that of HSV-1 strain 17 and one greater than that of strain KOS. The changed nucleotides were scattered and few, so that the first portion of the two leader sequences was 92% homologous. The next 49 bases were entirely different between the two types, and the HSV-1 leader from both strains was 16 bases longer than that in HSV-2. The entire block in Fig. 2 is shown as being different, because no significant homology could be detected, and the exact alignment varied depending on where the additional HSV-1 nucleotides were inserted. These changes are shown diagrammatically in Fig. 3.

Sequences further upsteam of the presumptive CAT homology show few intertypic changes. Within the 280 nucleotides that can be compared between the two upstream of nucleotide 320, 36 intertypic differences occurred, resulting in 87% homology. The 5.0-kb message that overlies the 1.2-kb message encodes a 140K protein in both HSV-1 (1) and HSV-2 (9). The complete sequence encoding this message has not yet been reported; therefore, the coding region has not been identified. In the phase that possesses an open reading frame for the 38K protein, there are five termination codons upstream of the start of translation, indicating that the coding sequences for the 140K and 38K proteins do not overlap, as has been shown for the HSV-1 polypeptides (6, 18). The other two reading frames each show one termination codon. In one case the stop codon TAA, located at positions 492 through 494, creates an open reading frame throughout the sequence extending to the *Bgl*II site and probably represents 163 amino acids of the carboxy terminus of the 140K protein. In HSV-1 strain 17, the same situation was found in that at the homologous position to the HSV-2 TAA the stop codon TGA occurred, leaving a long open reading frame. If these sequences actually represent the carboxy termini for the HSV-1 and HSV-2 140 K proteins, it is interesting that the coding sequences for one protein overlap the 5′ untranslated leader of another protein. Remarkably, immediately following the putative termination codon, the sequence of the HSV-2 DNA is drastically different from that of HSV-1 (see below).

Coding region. Beginning at the putative initiation codon (nucleotides 547 through 549) there was an open reading frame of 1,011 nucleotides that could encode a 337-amino-acid polypeptide. The termination codon TGA (nucleotides 1,558 through 1,560) was identical for HSV-1 and HSV-2. The sequence around the amino terminus of the protein differed greatly between HSV-1 and HSV-2. A block of 16 nucleotides surrounding the ATG, including 3 untranslated nucleotides, showed good homology with only two differences. Following that small stretch of homology, HSV-1 had an additional nine nucleotides compared with HSV-2. The next 10 nucleotides were identical in HSV-1 and HSV-2, followed by a stretch of 24 nucleotides which showed little homology. The next 42 nucleotides showed patchy homology with 14 changes, or only 66% homology. The sequence divergence had a striking effect on the first 30 amino acids (Fig. 4). Of the first eight amino acids of the HSV-2 protein, six were identical, and the other two changes were conservative compared with the HSV-1 protein. However, three additional amino acids, proline, alanine, and leucine, were inserted in the HSV-1-specified protein. Of the next 22 amino acids specified by the HSV-2 protein, there were 14 differences in the HSV-1 protein, 12 of which were nonconservative changes.

The remaining 307 amino acids of the protein were extremely similar between HSV-1 and HSV-2 with a total of 19 changes, only 10 of which were nonconservative. In general the changes were scattered throughout the protein except for a cluster from amino acids 240 to 253, which showed four nonconservative and two conservative changes out of the 14 amino acids. The overall relatedness of the HSV-1 strain KOS and HSV-2 38K proteins was 89%; excluding the amino terminus, it was 94%. The nucleotide homology was also quite good throughout the coding region with overall homology of 88%; excluding the sequences coding for the first 30 amino acids, it was 91%.

3′ region. The sequences that followed the coding region for the 38K protein were very different between HSV-1 and

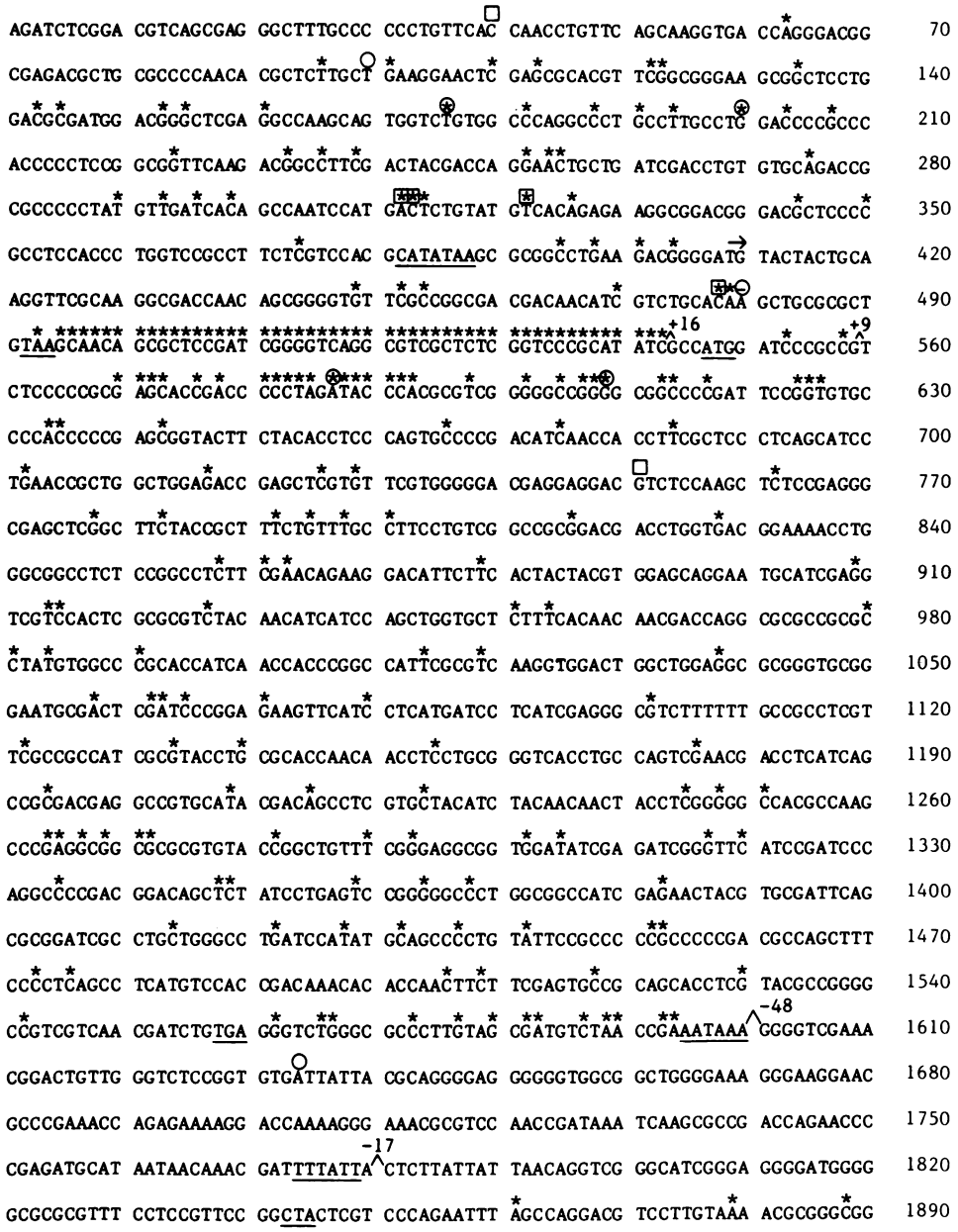

FIG. 2. Nucleotide sequence of the left-hand end of the *Bgl*II N fragment. The noncoding strand of the nucleotide sequence for the 38K protein is shown. Features of interest are underlined. Symbols above a nucleotide indicate the following: *, the nucleotide differed from the one that was present in both strains of HSV-1; −, the nucleotide was deleted in the HSV-1 sequence; +*n*, a number (*n*) of nucleotides were present in the HSV-1 sequence that were not present in the HSV-2 sequence; −*n*, a number (*n*) of nucleotides were deleted in the HSV-1 sequence; [*], a difference with HSV-1 strain KOS only; (*), a difference with HSV-1 strain 17 only; □ and ○, the extent of published sequence for HSV-1 strain KOS and HSV-1 strain 17, respectively. The arrow above nucleotide 410 indicates the start of the HSV-2 38K mRNA.

HSV-2. The canonical sequence signaling polyadenylation, AATAAA, was recognized (nucleotides 1,595 through 1,600) 35 bases following the stop codon. Thirteen of the 34 nucleotides separating the two signals differed between HSV-1 and HSV-2. Located 173 nucleotides downstream on the other strand was another polyadenylation signal (nucleotides 1,774 through 1,779) followed by 63 nucleotides to a potential stop codon (nucleotides 1,843 through 1,845). As part of a larger study (R. J. Frink, K. G. Draper, D. A. Galloway, and E. K. Wagner, unpublished results), we have tentatively identified these signals as belonging to the message that encodes the previously described HSV-1 58K protein (7) and the HSV-2 61K protein (9). The sequence that separated the two polyadenylation signals had undergone a great deal of divergence between HSV-1 and HSV-2. There were 48 fewer nucleotides in the HSV-1 sequence. Of the other 125 nucleotides, only very patchy homology could be detected. The longest stretch of detectable homology was 14 nucleotides, 12 of which were perfectly matched. In general the sequence could be aligned with stretches of three to six matched nucleotides, separated by stretches of similar length of changed or deleted nucleotides. The placement of

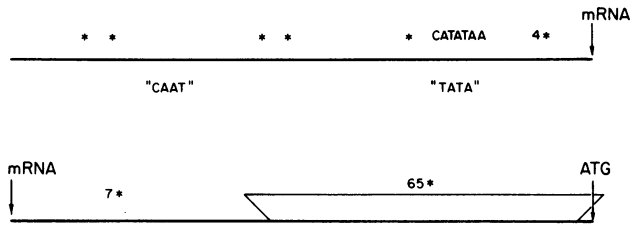

FIG. 3. Intertypic homology at the 5' end of the gene encoding 38K. The top line represents 135 nucleotides preceding the start of the RNA. Below the line are the predicted position of the CAAT and TATA homologies, and on the top line is the actual TATA sequence. The * represents changed nucleotides between the HSV-1 and HSV-2 sequence. The bottom line represents the untranslated leader RNA and indicates that the HSV-1 leader is longer than the HSV-2 leader, and that the entire block of sequences is different.

the deletions to create regions of homology was fairly arbitrary, so it is not shown in Fig. 2. Similarly, the DNA sequence separating the termination codon and the polyadenylation signal for the message encoding the protein on the other strand was also highly divergent between HSV-1 and HSV-2; 17 fewer nucleotides were present in that region of the HSV-1 sequence, and only patchy homology could be detected in the remaining 46 nucleotides. Once back in the sequences encoding a viral polypeptide, the homology became quite strong (data not shown, except for nucleotides 1,846 through 1,890).

DISCUSSION

The information derived from comparative DNA sequence analysis has provided interesting information about a virally encoded HSV-2-specific protein, the organization of coding and intergenic sequences in HSV, the genetic relatedness of HSV-1 and HSV-2, and some information about the mechanism of evolution of these viruses.

Despite the apparent type-specific nature of the HSV-2-encoded 38K protein, the homologous HSV-1 polypeptide is very similar. The overall protein homology is 89%; if the N-terminal 30 amino acids are considered separately, the rest of the protein is 94% homologous to its HSV-1 counterpart. In comparison, the thymidine kinase proteins of HSV-1 and HSV-2, which share antigenic determinants, show 73% homology. One interpretation is that the antigenicity of the 38K protein resides in the amino terminus of the molecule. Alternatively, the other scattered amino acid changes could be within antigenic sites, or the severe changes in the amino terminus caused by both amino substitutions and additions in the number of amino acids in the HSV-1 protein could alter the tertiary folding and hence the antigenicity. It is interesting that the type-specific epitopes of the 38K protein with which the monoclonal antibodies 6-H11 and 6-A6 react are highly conserved among independent isolates of HSV-2. In one study (14) these antibodies reacted with 187 isolates of HSV-2 and with none of 76 isolates that reacted with HSV-1-specific antibodies.

The relationship between the 38K and the 140K proteins remains obscure. Both immunoprecipitation (9, 14) and immunoblot (14) assays have suggested that the proteins share at least one related antigenic determinant. In addition, polyclonal sera specific for the 38K protein also react with the 140K protein (15), and antisera raised against the viral ribonucleotide reductase activity also precipitate both proteins. DNA sequence analysis has shown that the two proteins do not overlap in their coding sequences, as there were stop codons in all three phases upstream of the sequences encoding 38K. Although binding of the 140K and 38K proteins to each could explain the coprecipitation results, both the immunoblot data (14) and experiments in which the antibody 6-A6 reacted with the 38K protein, but not an altered 140K protein (K. Shriver, manuscript in preparation), argue for a common epitope. Whether the shared antigenic determinants represent functional domains of the proteins that are related, or merely a fortuitous repetition of some sequence, is unclear.

Analysis of this region of the viral genome has provided further evidence of how compact the genetic information is along the HSV genome. The DNA that presumably encodes the carboxy terminus of the 140K protein falls within sequences that code for the transcribed leader sequence of the message specifying the 38K protein. This feature is probably shared by both HSV-1 and HSV-2 and has not been previously described for other herpesvirus genes. The gene encoding the 61K protein, transcribed from the other strand, is not as closely spaced; in HSV-2 the putative termination codons are 282 nucleotides apart, and the polyadenylation signals are 173 nucleotides apart.

In the case of both the thymidine kinase gene and the gene specifying the 38K protein, the sequences at the 3' end of the gene showed the greatest intertypic divergence. The 3' region of the TK gene is 76% homologous and differs in number by two nucleotides, whereas the coding and 5' regions are 81% and 89% homologous, respectively (27). The divergence at the 3' end of the 38K gene was even more drastic. In the intergenic region containing the 3' ends of both the 38K and 61K genes there is virtually no homology, except for the polyadenylation signals and a short stretch following the termination codon for the 38K protein. It is difficult to comment on the degree of relatedness of the 5' region because the sequences serve not only as the 5' regulatory region for the 38K protein, but also as the apparent coding region and 3' end of the 140K protein. It is noteworthy that those sequences immediately following the

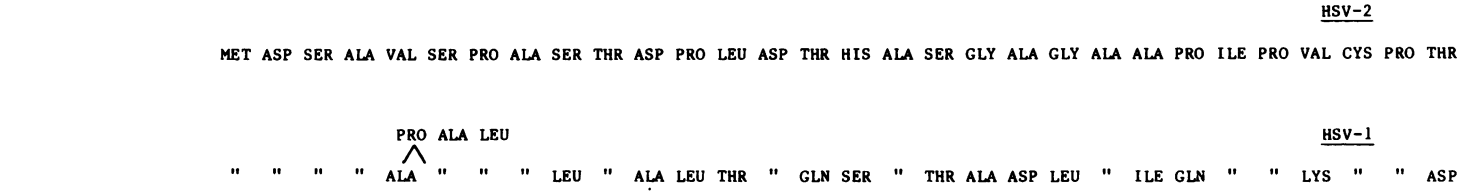

FIG. 4. Amino-terminal 30 residues of the 38K protein. The amino acid sequence predicted from the DNA sequence is shown. To obtain maximal homology, three additional amino acids were placed after residue 5 in HSV-1 protein. The symbol " indicates that the HSV-1 amino acid was the same as the HSV-2 amino acid.

potential termination codon for the 140K protein are highly divergent between HSV-1 and HSV-2; the number of nucleotides is not constant either, a situation analogous to that of the 3′ region for the 38K, 61K, and thymidine kinase proteins. The sequences that specify the transcriptional regulatory sequences (TATA) for the 38K protein and the first 80 nucleotides of the leader RNA are well conserved between HSV-1 and HSV-2. However, these sequences presumably also encode the carboxy terminus of the 140K protein, so it is unclear whether the constraints are directed toward the regulatory sequences or whether the constraints are on the protein-coding sequences. It is interesting that the leader sequence of the mRNAs is so different between HSV-1 and HSV-2; however it is not known whether this difference confers any biological phenotype on the 38K protein or on the stability of the message.

The basis of variation between the HSV-1 and HSV-2 38K genes is complex. Within the region we have sequenced, most of the variation in the protein-coding sequences (nucleotides 1 through 490 and 601 through 1560) was due to a series of single-base changes, as is the case for the entire region surrounding the TK gene. In the intergenic regions whole blocks of sequences have changed, including insertions and deletions. The block of sequences from nucleotides 495 through 543 shows no obvious homology with the 65 nucleotides from HSV-1, so it is possible that the entire segment represents a translocation. The sequences at the 3′ end of the gene show patchy homology, with very short stretches of homology separated by deletions of HSV-1 sequences and base changes. The DNA encoding the amino terminus of the 38K protein showed several types of variation. Within the fifth triplet, the HSV-1 sequence had an insertion of nine nucleotides that inserted additional amino acids into protein. In addition a stretch of 24 nucleotides showed no intertypic homology and the following 42 nucleotides showed no more variation (14 of 42) than occurred throughout the rest of the protein. Of the 1,850 nucleotides for which there are intertypic sequence data, the overall homology was 75%, which was somewhat lower than the 82% observed for the 1,600 nucleotides surrounding the TK gene, and the variation in the number of nucleotides is much greater. The overall intratypic variation within the 600 nucleotides that were compared between HSV-1 strain KOS and strain 17 showed only 1.5% variation, as did the region encoding the thymidine kinase gene from two other HSV-1 strains.

In summary, we have identified the sequences encoding the HSV-2 38K protein and presumably those encoding the carboxy termini of the 140K and 61K proteins. The 38K proteins of HSV-1 and HSV-2 are extremely similar, with the only significant change in the amino terminal 30 residues. The coding regions show the most intertypic homology, and the 3′ ends of the genes show the least conservation. HSV-1 and HSV-2 not only appear to have diverged as a consequence of scattered single base changes, but also show whole blocks of sequences that have changed, as well as the insertion or deletion of blocks of nucleotides.

ACKNOWLEDGMENTS

We thank Ed Wagner for sharing sequence data before publication. The HSV-1 sequence that we used for comparison is available from him in an amended form from the published (reference 6) version. We also acknowledge Steve Weinheimer for constructing the deletion fragments and sequencing them; Richard Gelinas for advice on M13 cloning, sequencing, and computer analysis; Devin Harrison for excellent assistance; Paul Su for artwork; and Ann Kritzberger for preparing the manuscript.

The work was supported by Public Health Sevice grants CA26001, CA35568 from the National Cancer Institute and by grants to the Fred Hutchinson Cancer Research Center.

LITERATURE CITED

1. **Anderson, K. P., R. J. Frink, G. B. Devi, B. H. Gaylord, R. H. Costa, and E. K. Wagner.** 1981. Detailed characterization of the mRNA mapping in the *Hin*dIII fragment K region of the herpes simplex virus type 1 genome. J. Virol. **37**:1011–1027.
2. **Aurelian, L., B. C. Strnad, and M. F. Smith.** 1977. Immunodiagnostic potential of virus-coded tumor associated antigen (AG-4) in cervical cancer. Cancer **39**:1834–1849.
3. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
4. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger and P. Chambon.** 1980. Promoter sequences of eukaryotic protein coding genes. Science **209**:1406–1415.
5. **Docherty, J. L., J. H. Subak-Sharpe, and C. M. Preston.** 1981. Identification of virus-specific polypeptide associated with a transforming fragment (*Bgl*II-N) of a herpes simplex virus type 2 DNA. J. Virol. **40**:126–132.
6. **Draper, K. G., R. J. Frink, and E. K. Wagner.** 1982. Detailed characterization of an apparently unspliced herpes simplex virus type 1 gene mapping in the interior of another. J. Virol. **43**:1123–1128.
7. **Frink, R. J., K. P. Anderson, and E. K. Wagner.** 1981. Herpes simplex virus type 1 *Hin*dIII fragment L encodes spliced and complementary mRNA species. J. Virol. **39**:559–572.
8. **Frink, R. J., K. G. Draper, and E. K. Wagner.** 1981. Uninfected cell polymerase efficiently transcribes early but not late herpes simplex virus type 1 mRNA. Proc. Natl. Acad. Sci. U.S.A. **78**:6139–6143.
9. **Galloway, D. A., L. C. Goldstein, and J. B. Lewis.** 1982. Identification of proteins encoded by a fragment of herpes simplex virus type 2 DNA that has transforming activity. J. Virol. **42**:530–537.
10. **Galloway, D. A., and J. K. McDougall.** 1981. Transformation of rodent cells by a cloned DNA fragment of herpes simplex virus type 2. J. Virol. **38**:749–760.
11. **Galloway, D. A., and J. K. McDougall.** 1983. The oncogenic potential of herpes simplex viruses: evidence for a "hit-and-run" mechanism. Nature (London) **302**:21–24.
12. **Gilman, S. C., J. J. Docherty, A. Clarke, and W. E. Rawls.** 1980. Reaction patterns of herpes simplex virus type 1 and type 2 proteins with sera of patients with uterine cervical carcinoma and matched controls. Cancer Res. **40**:4640.
13. **Gingeras, T. R., J. P. Milazzo, D. Sciaky, and R. J. Roberts.** 1979. Computer programs for the assembly of DNA sequences. Nucleic Acids Res. **7**:529–545.
14. **Goldstein, L. C., L. Corey, J. K. McDougall, E. Tolentino, and R. C. Nowinski.** 1983. Monoclonal antibodies to herpes simplex viruses: use in antigenic typing and rapid diagnosis. J. Infect. Dis. **147**:829–837.
15. **Hussar, D., and S. Bacchetti.** 1983. Is ribonucleotide reductase the transforming function of herpes simplex virus 2? Nature (London) **302**:76–79.
16. **Jariwalla, R. J., L. Aurelian, and P. O. P. Ts'O.** 1980. Tumorigenic transformation induced by a specific fragment of DNA from herpes simplex virus type 2. Proc. Natl. Acad. Sci. U.S.A. **77**:2279–2283.
17. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base specific chemical cleavage methods. Enzymology **65**:499–560.
18. **McLaughlan, J., and J. B. Clements.** 1983. Organization of the herpes simplex virus type 1 transcription unit encoding two early proteins with molecular weights of 140,000 and 40,000. J. Gen. Virol. **64**:997–1006.

18a. **McLaughlan, J., and J. B. Clements.** 1983. DNA sequence homology between two co-linear loci on the HSV genome which have different transforming activities. Eur. Mol. Biol. J. **2**:1953–

1961.
19. **Reyes, G. R., R. LaFemina, S. D. Hayward, and G. S. Hayward.** 1979. Morphological transformation by DNA fragments of human herpes viruses: evidence for two distinct transforming regions in herpes simplex virus types 1 and 2 and lack of correlation with biochemical transfer of the thymidine kinase gene. Cold Spring Harbor Symp. Quant. Biol. **44**:629–641.
20. **Sakonju, S., D. E. Bogenhagen, and D. D. Brown.** 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription. I. The 5′ border of the region. Cell **19**:13–25.
21. **Sanger, F., and A. R. Coulson.** 1975. A rapid method for determining sequences in DNA primed by synthesis with DNA polymerase. J. Mol. Biol. **94**:441–448.
22. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
23. **Schrier, P. H., and R. Cortese.** 1979. A fast simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. **129**:169–172.
24. **Staden, R.** 1979. Sequence data handling by computer. Nucleic Acids Res. **4**:4037–4051.
25. **Staden, R.** 1978. Further procedures for sequence analysis by computer. Nucleic Acids Res. **5**:1013–1015.
26. **Suh, M.** 1982. Characterization of a polypeptide present in herpes simplex virus type 2 transformed and infected hamster embryo cells. J. Virol. **41**:1095–1098.
27. **Swain, M. A., and D. A. Galloway.** 1983. Nucleotide sequence of the herpes simplex virus type 2 thymidine kinase gene. J. Virol. **46**:1045–1050.