JOURNAL OF VIROLOGY, Nov. 1984, p. 715–718
0022-538X/84/110715-04$02.00/0

Vol. 52, No. 2

A Tandemly Reiterated DNA Sequence in the Long Repeat Region of Herpes Simplex Virus Type 1 Found in Close Proximity to Immediate-Early mRNA 1

FRAZER J. RIXON, MOYRA E. CAMPBELL, AND J. BARKLIE CLEMENTS*

Institute of Virology, University of Glasgow, Glasgow G11 5JR, Scotland

Received 12 April 1984/Accepted 30 July 1984

The 3′ end of immediate-early mRNA 1 was mapped precisely within the IR_L/TR_L genome regions, and the DNA sequences around the 3′ end were determined. An AATAAA polyadenylation signal was present 17 base pairs upstream of the 3′ end, and eight tandemly repeated copies of a 16-base-pair sequence (GGGGGTGCGTGGGAGT) plus one further closely related copy were located 20 base pairs downstream. Other tandem reiterations present in the herpes simplex virus genome are described and their properties are considered.

The herpes simplex virus (HSV) genome contains a number of different short, tandemly reiterated DNA sequences. These tandem reiterations, although individually distinct in sequence, share certain common features. They are of short length, ranging from 5 base pairs to 37 (bp), have high guanine plus cytosine contents, show marked asymmetry in the distribution of purines and pyrimidines on the two DNA strands, and their copy numbers vary between individual virus genomes. The variation in copy number appears as size heterogeneity of those bands which contain the reiterations in the profile of DNA fragments generated by restriction endonucleases (9). When these cloned DNA fragments are sequenced, the variation in copy number becomes evident (3, 14, 16, 27).

So far, tandem reiterations have been identified directly by DNA sequencing only within the short genome regions of both HSV type 1 (HSV-1) (3, 11, 14, 16, 19, 27; D. J. McGeoch, S. Donald, A. Dolan, and F. J. Rixon, submitted for publication) and HSV-2 (28) and are located in close proximity to those immediate-early (IE) genes which map in the IR_S and TR_S genome regions (3, 16, 27, 28).

We report here on the identification of a reiterated DNA sequence which is located within the long repeat regions (IR_L and TR_L) of HSV-1, just downstream from the 3′ end of IE mRNA 1 (see Fig. 2).

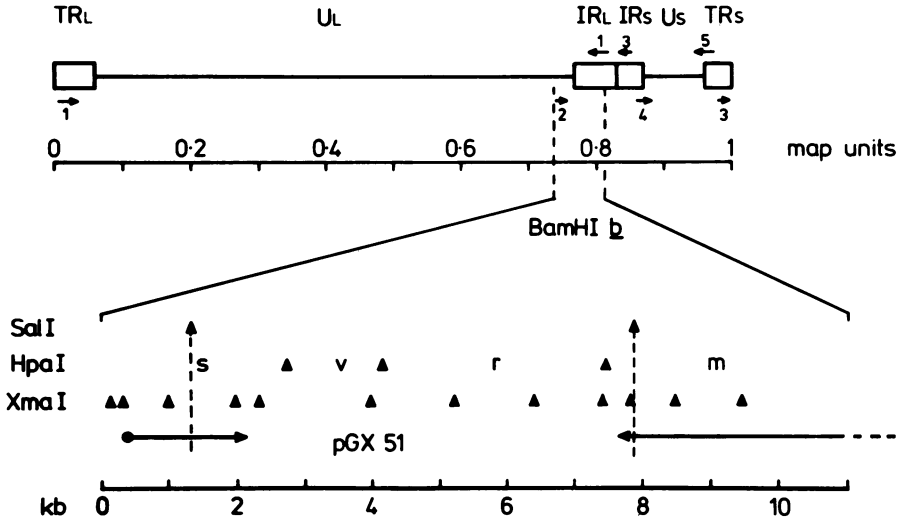


FIG. 1. HSV-1 genome shown in the prototype orientation (15), together with the map locations and orientations of the five major IE mRNAs (2). The long and short regions of the genome consist of unique regions (U_L and U_S) which are flanked and joined by inverted, repeated regions (TR_L/IR_L and TR_S/IR_S, respectively) (22). The restriction endonuclease cleavage sites and cloned DNA fragments used for nuclease S1 mapping of IE mRNA 2 are shown. *Hpa*I fragments s, v, r, and m are indicated. The *Xma*I map is incomplete.

The 3′ end of IE mRNA 1 was located by using plasmid pGX51, which contains the *Sal*I subfragment of *Bam*HI-b (Fig. 1). Plasmid pGX51 DNA was 3′ labeled at the two *Sal*I sites and then hybridized to cytoplasmic IE mRNA and to mock-infected mRNA samples as described previously (18). Two nuclease S1-resistant products of 790 and 240 bases were observed; the 790-base product was formed by the 3′ portion of IE mRNA 2 (Fig. 1), whereas the 240-base

* Corresponding author.

product was formed by the 3′ end of IE mRNA 1 (29).

To position precisely the 3′ end of IE mRNA 1 on the DNA sequence, pGX51, 3′ labeled at the *Sal*I sites, was further digested with *Hpa*I (Fig. 1). A portion of the uniquely labeled 440-bp *Sal*I-*Hpa*I fragment, which encodes the 3′ end of IE mRNA 1, was used for nuclease S1 analysis, and the remainder was sequenced by the procedure of Maxam and Gilbert (10). The products of these reactions were electrophoresed on an 8% denaturing polyacrylamide gel (Fig. 2) which revealed a highly reiterated DNA sequence located just downstream from the 3′ end. To determine this sequence, a 450-bp *Xma*I fragment (Fig. 1) containing the reiteration was sequenced; the sequence (Fig. 3) was confirmed by M13-dideoxy sequencing (13, 20) of the equivalent *Sma*I fragment (*Sma*I is an isoschizomer of *Xma*I).

The major nuclease S1-resistant band of 240 bases places the 3′ end at position 230 (Fig. 3), ca. 17 bp downstream from an AATAAA polyadenylation signal (4). The sequence TGTGTTGG, which is present 12 to 19 bp downstream from the 3′ end, corresponds to sequences (consensus YGTGTTYY) located ca. 15 bp downstream from the 3′ termini of a number of HSV mRNAs (12, 28, 29). A second AATAAA sequence located at position 170 and a GT-rich sequence (TCTGTTGG) at position 199 are in a similar spatial relationship to those at positions 209 and 242 (Fig. 3); however, no mRNA 3′ end was found associated with these sequences.

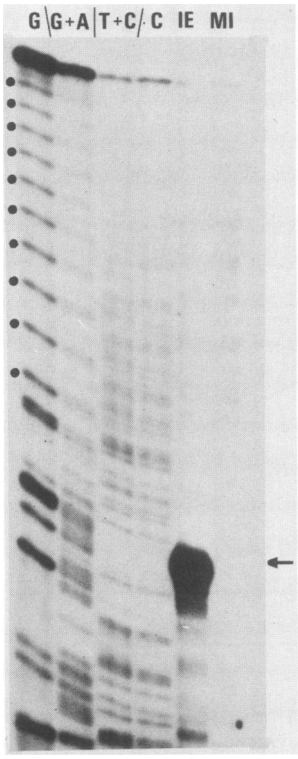


FIG. 2. Location of the 3′ end of IE mRNA 2. A 440-bp *Sal*I-*Hpa*I subfragment of pGX51 (Fig. 1) was 3′ labeled at the *Sal*I site; a portion of this material was sequenced, and the remainder was hybridized with either 20 μg of mock-infected cytoplasmic RNA (MI) or 15 μg of IE cytoplasmic RNA (IE). The nuclease S1-resistant material and sequence reaction products were electrophoresed on an 8% denaturing polyacrylamide gel. Dots alongside the G reaction track indicate the single G residue present in the nine complete and one incomplete copies of the reiteration (Fig. 3). The arrow indicates the nuclease S1-resistant band formed by the 3′ portion of IE mRNA 1. MI, Define.

The DNA sequence reiteration begins ca. 25 bp downstream from the 3′ end: eight complete copies and an incomplete copy of the sequence GGGGGTGCGTGGGAGT are preceded by an imperfect copy (GGGGATGCGTG GGAGT) which has an A instead of a G at position 5 (Fig. 3). This sequence reiteration present in TR_L/IR_L, together with the other tandemly reiterated DNA sequences identified in the HSV short genome region, are shown in Fig. 4.

Smith (24) has proposed that a DNA region which is not under selective pressure will become and remain, repetitious as a result of random, nonhomologous recombination. Tautz and Renz (26) have shown that simple sequences comprising one-, two-, and three-bp repeats are present in genomic DNAs from a wide spectrum of eucaryotes, and they suggested that these repeats have no specific function and probably arose through unequal crossover or replicative slippage. Tandem reiterations of longer sequences, resembling those present in HSV, also are found in eucaryotic genomes, both in polypeptide-coding regions, where they encode reiterated amino acid sequences (5, 6), and in noncoding regions (1, 7, 17, 23). Reiterations within polypeptide-coding sequences are likely to be under selective pressure and may represent a different class from those considered by Smith (24).

In the HSV genome, tandem reiterations occur in both polypeptide-coding (10, 19; McGeoch et al., manuscript in preparation) and noncoding regions (3, 11, 14, 16, 27, 28; McGeoch et al., manuscript in preparation). As certain non-polypeptide-coding reiterations in HSV-1 have no equivalent in HSV-2 (Fig. 4), this reinforces the view that these reiterations have no direct function. Although most of the HSV reiterations described so far comprise direct repeats of a single DNA sequence, a tandemly reiterated family with combinations of three different but related sequences has been identified within the intron of HSV-2 IE mRNAs 4 and 5 (Fig. 4) (28). These more complex reiterations resemble the higher order repeats which Smith (24) predicted would be generated by continued nonhomologous recombination involving simple sequence repeats.

Tandemly reiterated sequences which are not under selective pressure would show a tendency to expand and occupy nonessential portions of the genome (24). This represents the situation shown by the reiterations at location IV (Fig. 4) which, in both HSV-1 and HSV-2, occupy virtually the entire sequence of an intron apart from those sequences required for splicing (21), and also is shown by the HSV-1 reiterations at locations III (3), VII (McGeoch et al., manuscript in preparation), and VIII (Fig. 3) which extend to within 30 bp of mRNA 3′ termini.

In tandemly reiterated sequences, homologous recombination can occur between any members of the sequence family, and out-of-register crossover between imperfectly aligned families generates progeny molecules with differing copy numbers of reiterations (25). The marked copy number variability of the reiterated sequences present in the herpesvirus genome (3, 16) and in the eucaryotic genome (1, 8) indicates that these sequences are subject to high rates of recombination.

Intermolecular recombination involving HSV DNA molecules with relatively inverted orientations of the L and S regions will result in transfer of genetic information between IR_L and TR_L and between IR_S and TR_S. Thus, mutations

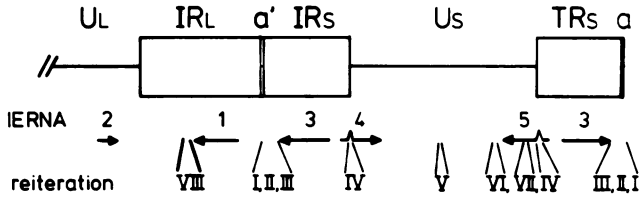
```
        10                  20                  30                  40                  50                  60
CCCGGGACGA          GGGAAAACAA          TAAGGGACGC          CCCCCGTGTT          TGTGGGGAGG          GGGGGGTCGG

        70                  80                  90                  100                 110                 120
GCGCTGGGTG          GTCTCTGGCC          GCGCCCACTA          CACCAGCCAA          TCCGTGTCGG          GGAGGGGAAA

        130                 140                 150                 160                 170                 180
AGTGAAAGAC          ACGGGCACCA          CACACCAGCG          GGTCTTTTGT          GTTGGCCCTA          ATAAAAAAAA

        190                 200                 210                 220                 230                 240
ACTCAGGGAT          TTTTGCTGTC          TGTTGGGAAA          TAAAGGTTTA          CTTTTGTATC          TTTTCCCTGT
                                                                                                    ●●●●
        250
CTGTGTTGGA          TGTATCGC    [GGGGATGCGTGGGAGT] [GGGGGTGCGTGGGAGT]     [GGGGGTGC]
                                                                                x8
        420                 430                 440                 450
CATGTTGGGC          AGGCTCTGGT          GTTAACCACA          GAGCCGCGGC          CCGGG
```

FIG. 3. Nucleotide sequence of a 450-bp *Xma*I fragment encoding the 3′ end of IE mRNA 2. The dots below the sequence, at position 230, indicate the position of the 3′ end. The two AATAAA sequences are underlined as is the sequence YGTGTTYY (Y, pyrimidine). The brackets enclose the two forms of the 16-bp DNA sequence present as nine direct repeats and a partial copy.



		bp
HSV-1		
I	CCGCTCCTCCCC[1]	12
	CCGCTCCTCCC[2]	11
	CCGCTCCTCCC[2]	24
	TCTGTGGGTGGGG	
	CCGCTCCTCCCC[3]	12
	CCGCTCCTCCCC[3]	37
	GCTCCCGCGGCCCCGCCCCCAACGC	
II	CCCTCCCCAGCCCCAG[1]	16
	CCCTCCCCGGCCCCAG[1]	16
III	CCGCCCCTCGCCCCCTC[1]	17
IV	CCCCCTCCTCCGCCCCCGCGTC[1]	22
	CCCCCTCCTCCACCCCCGCGTC[1]	22
	CCCCCTCCTCCGCCCCCGCGTCC[4]	23
V	CCTCCACCCCCTCGACCACCA[1]	21
VI	CTCCCCACCCACCCA[1]	15
VII	CCCCGGTCTCCCCGGGAG[1]	18
VIII	ACTCCCACGCACCCCC[1]	16
	ACTCCCACGCATCCCC[1]	16
HSV-2		
IV	GTCCCCCC	8
	GGCCCCCC	8
	GGCCC	5

FIG. 4. Summary of reiterated DNA sequences described so far in the HSV genome portion which comprises the entire short region, part of the long region including the IR_L inverted repeat, and the 'a' sequence at the joint, which is also present at each end of the genome. Locations of the different tandemly reiterated sequences are indicated as are the locations of the major IE mRNAs. Reiteration I lies within the a sequence; reiterations II and III lie between the a sequence and the 3′ end of IE mRNA 3; reiteration IV lies within the common intron of IE mRNAs 4 and 5; reiteration VI lies between the 3′ end of IE mRNA 5 and the 3′ end of the mRNA encoding glycoprotein E; reiterations V and VII lie within polypeptide-coding sequences; reiteration VIII is described here. The sequences of the individual reiterations are given below their genome locations. The HSV-1 sequences, as indicated by the superscript numbers, were derived from the following strains: 1, MP17 (3, 11, 16, 19; this paper); 2, USA-8 (3); 3, F (14); 4, Patton (27). In addition, strain F has a reiterated sequence at location II which is identical to one of those shown for MP17. All sequences show the pyrimidine-rich strand written in 5′→3′ orientation. Reiterations II, IV, and VIII in HSV-1 are present as two forms which differ from each other in a single residue (3, 16; this paper). For reiteration I in strain USA-8, the situation is more complex, as this family of reiterations contains two components: the first comprises 11 of the 12 bases reiterated in strain MP17, adjacent to which is an additional reiteration which contains the same 11 bases with a further 13 bases inserted as shown. Reiteration 1 in strain F has a similar arrangement with several tandem repeats of the 12-base sequence shown, adjacent to a reiteration which consists of the same 12 bases linked to a further 25 bases in the manner indicated. The HSV-2 strain HG52 reiterations at location IV comprise the three sequences shown, repeated in an irregular arrangement (28). A tandemly reiterated DNA sequence is known to be present at location VIII in HSV-2 strain HG52, but the precise sequence has not yet been determined (L. J. Whitton, personal communication); no HSV-2 tandem reiterations are present at locations I (3) or VI (L. J. Whitton, Ph.D. thesis, University of Glasgow, Glasgow, Scotland, 1984).

which arise in one copy of an inverted repeat will be transferred to the other copy and, in a replicating population of molecules, this will randomize the distribution of a mutation between both copies. Once a mutation is present in both copies of an inverted repeat, it will be exposed to normal selective pressures, thus facilitating the loss of disadvantageous mutations and promoting the maintainance of advantageous mutations. By promoting a high rate of recombination, the tandem reiterations should enhance this process of genetic exchange, thereby reducing the buildup of recessive mutations within the repeat regions.

We thank A. J. Davison for his advice and assistance and J. H. Subak-Sharpe for his support and for critical reading of this manuscript.

This work was supported by a Medical Research Council grant G978/709/5.

LITERATURE CITED

1. **Bell, G. I., M. J. Selby, and W. J. Rutter.** 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature (London) **295**:31–35.
2. **Clements, J. B., J. McLauchlan, and D. J. McGeoch.** 1979. Orientation of herpes simplex virus type 1 immediate-early mRNAs. Nucleic Acids Res. **7**:77–91.
3. **Davison, A. J., and N. M. Wilkie.** 1981. Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. J. Gen. Virol. **55**:315–331.
4. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24**:251–260.
5. **Garfinkle, M. D., R. E. Pruitt, and E. M. Meyerowitz.** 1983. DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. **168**:765–789.
6. **Godson, G. N., J. Ellis, P. Svec, D. H. Schlesinger, and V. Nussenzweig.** 1983. Identification and chemical synthesis of a tandemly repeated immunogenic region of Plasmodium knowlesi circumsporozoite protein. Nature (London) **305**:29–33.
7. **Heilig, R., R. Muraskowsky, and J.-L. Mandel.** 1982. The ovalbumin gene family. The 5′ end region of the X and Y genes. J. Mol. Biol. **156**:1–19.
8. **Lebo, R. V., A. Chakravarti, K. H. Buetow, M.-C. Cheung, H. Cann, B. Cordell, and H. Goodman.** 1983. Recombination within and between the human insulin and β-globin gene loci. Proc. Natl. Acad. Sci. U.S.A. **80**:4808–4812.
9. **Lonsdale, D. M., S. M. Brown, J. Lang, J. H. Subak-Sharpe, H. Koprowski, and K. G. Warren.** 1980. Variations in herpes simplex virus isolated from human ganglia and a study of clonal variation in HSV-1. Ann. N.Y. Acad. Sci. **354**:291–308.
10. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. **65**:499–560.
11. **McGeoch, D. J.** 1984. The nature of animal virus genetic material, p. 75–107. *In* B. W. J. Mahy and J. R. Pattison (ed.), The microbe—1984. Part I: viruses. Cambridge University Press, Cambridge, England.
12. **McLauchlan, J., and J. B. Clements.** 1980. DNA sequence homology between two co-linear loci on the HSV genome which have different transforming abilities. EMBO (Eur. Mol. Biol. Org.) J. **2**:1953–1961.
13. **Messing, J., and J. Vieira.** 1983. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
14. **Mocarski, E. S., and B. Roizman.** 1981. Site-specific inversion sequence of the herpes simplex virus genome: domain and structural features. Proc. Natl. Acad. Sci. U.S.A. **78**:7047–7051.
15. **Morse, L. S., T. G. Buchman, B. Roizman, and P. A. Schaffer.** 1977. Anatomy of herpes simplex virus DNA. IX. Apparent exclusion of some parental DNA arrangements in the generation of intertypic (HSV-1 × HSV-2) recombinants. J. Virol. **24**:231–248.
16. **Murchie, M.-J., and D. J. McGeoch.** 1982. DNA sequence analysis of an immediate-early gene region of the herpes simplex virus type 1 genome (map coordinates 0.950 to 0.978) J. Gen. Virol. **62**:1–15.
17. **Nikaido, T., S. Nakai, and T. Honjo.** 1981. Switch region of immunoglobulin Cμ gene is composed of simple tandem repetitive sequences. Nature (London) **292**:845–848.
18. **Rixon, F. J., and J. B. Clements.** 1982. Detailed structural analysis of two spliced HSV-1 immediate-early mRNAs. Nucleic Acids Res. **10**:2241–2256.
19. **Rixon, F. J., and D. J. McGeoch.** 1984. A 3′ co-terminal family of mRNAs mapping in the short region of HSV-1. Nucleic Acids Res. **12**:2473–2487.
20. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
21. **Seif, I., G. Khoury, and R. Dhar.** 1979. BKV splice sequences based on analysis of preferred donor and acceptor sites. Nucleic Acids Res. **6**:3387–3398.
22. **Sheldrick, P., and N. Berthelot.** 1974. Inverted repetitions in the chromosome of herpes simplex virus. Cold Spring Harbor Symp. Quant. Biol. **39**:667–678.
23. **Slightom, J. L., A. E. Blechi, and O. Smithies.** 1980. Human fetal Gγ- and Aγ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.
24. **Smith, G. P.** 1976. Evolution of repeated DNA sequences by unequal crossover. Science **191**:528–535.
25. **Sturtevant, A. H.** 1925. The effect of unequal crossing-over at the Bar locus in *Drosophila*. Genetics **10**:117–147.
26. **Tautz, D., and M. Renz.** 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. **12**:4127–4138.
27. **Watson, R. J., K. Umene, and L. W. Enquist.** 1981. Reiterated sequences within the intron of an immediate-early gene of herpes simplex virus type 1. Nucleic Acids Res. **9**:4189–4199.
28. **Whitton, J. L., and J. B. Clements.** 1984. The junctions between the repetitive and the short unique sequences of the herpes simplex virus genome are determined by the polypeptide coding regions of two spliced immediate-early mRNAs. J. Gen. Virol. **65**:451–466.
29. **Whitton, J. L., F. J. Rixon, A. J. Easton, and J. B. Clements.** 1983. Immediate-early mRNA-2 of herpes simplex viruses types 1 and 2 is unspliced: conserved sequences around the 5′ and 3′ termini correspond to transcription regulatory signals. Nucleic Acids Res. **11**:6271–6287.