Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 1846–1849, March 1989
Biochemistry

Complete nucleotide sequence of a molecular clone of woodchuck hepatitis virus that is infectious in the natural host

(recombinant DNA/liver transfection)

ROSINA GIRONES*, PAUL J. COTE†, WILLIAM E. HORNBUCKLE‡, BUD C. TENNANT,‡, JOHN L. GERIN†, ROBERT H. PURCELL*, AND ROGER H. MILLER*

*Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892; †Georgetown University, Division of Molecular Virology and Immunology, Rockville, MD 20852; and ‡Cornell University, College of Veterinary Medicine, Ithaca, NY 14853

Contributed by Robert H. Purcell, December 23, 1988

ABSTRACT **Woodchuck hepatitis virus (WHV) DNA was cloned from viral particles obtained from the serum of a woodchuck with a naturally acquired infection. The complete nucleotide sequence of the virus genome was determined and found to be 3323 base pairs long. Transfection experiments demonstrated that the recombinant WHV DNA was infectious in each of 18 woodchucks tested and established a chronic carrier state in 1 of 13 neonates and 3 of 5 adult animals. WHV DNA from serum particles from the chronically infected neonate was cloned and the nucleotide sequence of three independent recombinants was compared directly with that of the input recombinant DNA. The consensus sequence of the three progeny genomes was identical to that of the parental DNA sequence. Therefore, transfection of woodchuck livers with recombinant WHV DNA induces active virus replication and gene expression and yields progeny genomes that are faithful copies of the input virus genome.**

Hepatitis B virus (HBV) is the prototype of the hepadnavirus family, which consists of hepatotropic viruses that infect at least six separate species. Infection with HBV causes polymorphic liver diseases, including acute or fulminant hepatitis, chronic hepatitis, cirrhosis, and hepatocellular carcinoma. Approximately 300 million people are chronically infected with HBV and are predisposed to developing hepatocellular carcinoma, one of the world's most prevalent forms of cancer.

The study of woodchucks infected with woodchuck hepatitis virus (WHV) provides an animal model system for the study of the biological and pathological properties of HBV. The WHV genome shares ≈65% homology with the HBV genome (1) and the woodchuck hepatitis surface antigen (WHsAg) shares antigenic cross-reactivity with the HBV surface antigen. Also, the risk of hepatocellular carcinoma in experimentally infected chronic WHsAg carriers is 100% (2). Thus, WHV infection of woodchucks appears to be a valuable experimental system for the study of hepadnavirus replication, gene expression, and the development of hepatocellular carcinoma.

WHV has been isolated from numerous woodchucks (*Marmota monax*) from the eastern United States. The genomes of several of these isolates have been cloned and the nucleotide sequence determined: WHV1 (1), WHV2 (3), WHV7, and WHV59 (4). We present here the nucleotide sequence of a fifth WHV clonal isolate, WHV8, and demonstrate that the recombinant DNA is infectious by DNA transfection of the livers of neonatal and adult woodchucks.§

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

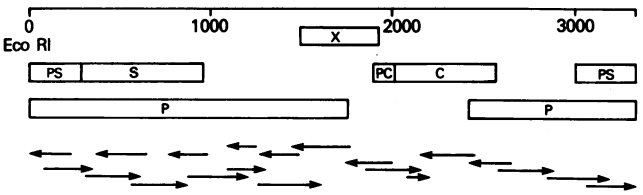

FIG. 1. Strategy of sequence analysis of pWHV8 DNA. The four known WHV genes are depicted as boxes: C, core; PC, precore; S, surface; PS, presurface; P, polymerase; and X. The nucleotide sequence was determined in the regions of the genome underscored by the arrows.

MATERIALS AND METHODS

Cloned DNA. WHV8 DNA was isolated from serum particles from a chronically infected woodchuck, woodchuck number 8 (WC8). This naturally infected animal was captured in northeastern Maryland in 1978 and died with primary hepatocellular carcinoma in 1980. The WHV DNA was cloned into the *Eco*RI site of pBR325 by standard methods (5). The recombinant was designated pWHV8.

Transfection Procedures. pWHV8 DNA was amplified in *Escherichia coli* and plasmid DNA was purified by cesium chloride/ethidium bromide buoyant density centrifugation. WHV8 DNA was separated from vector DNA sequences by digestion with the restriction endonuclease *Eco*RI followed by agarose gel electrophoresis and electroelution. The WHV8 genome was then ligated, using a DNA concentration (i.e., 10 μg/ml) favoring the production of monomeric circular molecules. Ligated DNA was precipitated and resuspended in phosphate-buffered saline (PBS). The livers of neonatal (3–7 days old) woodchucks were transfected with 10 μg of WHV8 DNA in 50 μl of PBS by percutaneous inoculation as described for polyomavirus transfection of neonatal mice (6). Adult woodchuck livers were surgically exposed and transfected in six separate sites with a total of 40 μg of WHV8 DNA in 100–200 μl of PBS or as a calcium phosphate precipitate analogous to the methods described for ground squirrel hepatitis virus DNA transfection of ground squirrel livers (7).

Virus particles were isolated from the serum of a chronically infected woodchuck (WC1806) 16 months after neonatal transfection. The circular partially double-stranded genome was made double stranded by the endogenous DNA polymerase reaction and virus DNA was then purified from virions. Any remaining single-stranded DNA was converted

Abbreviations: HBV, hepatitis B virus; WHV, woodchuck hepatitis virus; WHsAg, woodchuck hepatitis surface antigen; anti-WHs, woodchuck hepatitis surface antibody; anti-WHc, woodchuck hepatitis core antibody.

§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04514).

to double-stranded DNA with avian myeloblastosis virus reverse transcriptase, in the presence of all four dNTPs and Mg^{2+}, and the resulting fully double-stranded WHV8′ DNA was cloned into the *Hin*dIII site of pUC13. Three independent recombinants (pWHV8′ a, b, and c) were selected for sequence analysis.

Nucleotide Sequence. The complete nucleotide sequence of pWHV8 and three progeny genomes pWHV8′ (clones a, b, and c) was determined by the dideoxynucleotide sequencing technique using oligonucleotide primers (4, 8). Ambiguities in the sequences were resolved by sequencing both DNA strands in overlapping regions of the recombinant genome. The sequencing strategy for pWHV8 DNA is shown in Fig. 1.

Serological Analysis. Serum samples were obtained from all animals at monthly intervals for the first 9 months and at quarterly intervals thereafter. Animals were followed for ≈2

```
                                                                                                                    100
AATTCGGGACATACCACGTGGTTTAGTTCCGCCTCAAACTCCAACAAATCGAGATCAAGGGAGAAAGCCTACTCCTCCAACTCCACCTCTAAGAGATACT
                                                                                                                    200
CACCCCCACTTAACTATGAAAAATCAGACTTTTCATCTCCAGGGGTTCGTAGACGGATTACGAGACTTGACAACAACGGAACGCCAACACAATGCCTATG
                                                                                                PRE-SURFACE ENDS>< 300
GAGATCCTTTTACAACACTAAGCCCTGCGGTTCCTACTGTATCCACCATATTGTCTCCTCCCTCGACGACTGGGGACCCTGCACTGTCACCGGAGATGTC
SURFACE STARTS                                                                                                      400
ACCATCAAGTCTCCTAGGACTCCTCGCAGGATTACAGGTGGTGTATTTCTTGTGGACAAAAATCCTAACAATAGCTCAGAATCTAGATTGGTGGTGGACT
                                                                                                                    500
TCTCTCAGTTTTCCAGGGGGCATACCAGAGTGCACTGGCCAAAATTCGCAGTTCCAAACTTGCAAACACTTGCCAACCTCCTGTCCACCAACTTGCAATG
                                                                                                                    600
GCTTTCGTTGGATGTATCTGCGGCGTTTTATCATATACCTATTAGTCCTGCTGCTGTGCCTCATCTTCTTGTTGGTTCTCCTGGACTGGAAAGGTTTAAT
                                                                                                                    700
ACCTGTCTGTCCTCTTCAACCCACAACAGAAACAACAGTCAATTGCAGACAATGCACAATCTCTGCACAAGACATGTATACTCCTCCTTACTGTTGTTGT
                                                                                                                    800
TTAAAACCTACGGCAGGAAATTGCACTTGTTGGCCCATCCCTTCATCATGGGCTTTAGGAAATTACCTATGGGAGTGGGCCTTAGCCCGTTTCTCTTGGC
                                                                                                                    900
TCAATTTACTAGTGCCCTTGCTTCAATGGTTAGGAGGAATTTCCCTCATTGCGTGGTTTTTGCTTATATGGATGATTTGGTTTTGGGGGCCCGCACTTCT
                                                                                    SURFACE STOPS >                1000
GAGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTGGGTATACATTTGAATGTCAATAAAACAAAATGGTGGGGCAATCATCTAC
                                                                                                                   1100
ATTTCATGGGATATGTGATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAAATTTCCCGTTATTTGCACTCTGTTCCTGTTAATCAACC
                                                                                                                   1200
TCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
                                                                                                                   1300
ATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGG
                                                                                                                   1400
TGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCAC
                                                                                                                   1500
GGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAGCTGACGTCCTTT
  < X STARTS                                                                                                       1600
CCATGGCTGCTCGCCTGTGTTGCCCACTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCC
                                                                                                                   1700
TGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGTTTCGCCTCGGCGTCC
                  < D.R.   >        POLYMERASE STOPS >                                                             1800
GGTCCGTGTTGCTTGGTCTTCACCTGTGCAGACTTGCGAACCATGGATTCCACCGTGAACTTTGTCTCCTGGCATGCAAATCGTCAACTTGGCATGCCAA
                                                                                                                   1900
GTAAGGACCTTTGGACTCCTTATATAAAAGATCAATTATTAACTAAATGGGAGGAGGGCAGCATTGATCCTAGATTATCAATATTTGTATTAGGAGGCTG
                X STOPS >              <    D.R. >                                                                 2000
TAGGCATAAATGCATGCGACTTCTGTAACCATGTATCTTTTTCACCTGTGCCTTGTTTTTGCCTGTGTTCCATGTCCTACTTTTCAAGCCTCCAAGCTGT
                  < CORE STARTS                                                                                    2100
GCCTTGGATGGCTTTGGGGCATGGACATAGATCCCTATAAAGAATTTGGTTCATCTTATCAGTTGTTGAATTTTCTTCCTTTGGACTTCTTTCCTGACCT
                                                                                                                   2200
TAATGCTTTGGTGGACACTGCTACTGCCTTGTATGAAGAAGAGCTAACAGGTAGGGAACATTGCTCTCCGCACCATACAGCTATTAGACAAGCTTTAGTA
                                                                                                                   2300
TGCTGGGATGAATTAACTAAATTGATAGCTTGGATGAGCTCTAACATAACTTCTGAACAAGTAAGAACAATCATAGTAAATCATGTCAATGATACCTGGG
                                                                                                                   2400
GACTTAAGGTGAGACAAAGTTTATGGTTTCATTTGTCATGTCTCACTTTCGGACAACATACAGTTCAAGAATTTTTAGTAAGTTTTGGAGTATGGATCAG
                      < POLYMERASE STARTS                                                                          2500
AACTCCAGCTCCATATAGACCTCCTAATGCACCCATTCTCTCGACTCTTCCGGAACATACAGTCATTAGGAGAAGAGGAGGTGCAAGAGCTTCTAGGTCC
                                                                                          CORE STOPS  >            2600
CCCAGAAGACGCACTCCCTCTCCTCGCAGGAGAAGATCTCAATCACCGCGTCGCAGACGCTCTCAATCTCCATCTGCCAACTGCTGATCTTCAATGGGTA
                                                                                                                   2700
CATAAAACTAATGCAATTACAGGTCTTTACTCTAACCAAGCTGCTCAGTTCAATCCGAATTGGATTCAACCTGAGTTTCCTGAACTTCATTTACATAATG
                                                                                                                   2800
ATTTAATTCAAAAATTGCAACAGTATTTTGGTCCTTTGACTATAAATGAAAAGAGAAAATTGCAATTAAATTTTCCTGCCAGATTTTTCCCCAAAGCTAC
                                                                                                                   2900
TAAATATTTCCCTTTAATTAAAGGCATAAAAAACAATTATCCTAATTTTGCTTTAGAACATTTCTTTGCTACCGCAAATTATTTGTGGACTTTATGGGAA
                                                                                                               <   3000
GCTGGAATTTTGTATTTAAGGAAGAATCAAACAACTTTGACTTTTAAAGGTAAACCATATTCTTGGGAACACAGACAGCTAGTGCAACATAATGGGCAAC
PRE-SURFACE STARTS                                                                                                 3100
AACATAAAAGTCACCTTCAATCCAGACAAAATAGCAGCATGGTGGCCTGCAGTGGGCACTTATTACACAACCACTTATCCTCAGAATCAGTCAGTGTTTC
                                                                                                                   3200
AACCAGGAATTTATCAAACAACATCTCTGATAAATCCCAAAAATCAACAAGAACTGGACTCTGTTCTTATAAACAGATACAAACAGATAGACTGGAACAC
                                                                                                                   3300
TTGGCAAGGATTTCCTGTGGATCAAAAATTTTCATTGGTCAGCAGGGATCCTCCCCCAAAACCTTATATAAATCAATCAGCTCAAACTTTCGAAATCAAA
                    3323
CCTGGGCCTATAATAGTTCCCGG
```

FIG. 2. Complete nucleotide sequence of the WHV8 genome. The sequence of WHV8 DNA is presented with the *Eco*RI site as map position 0. The initiation (start) and termination (stop) codons of the four known WHV genes are shown above the nucleotide sequence. DR, direct repeat.

Table 1. Infectivity of recombinant WHV8 DNA

Transfection	Months observed	Number of woodchucks	Positive anti-WHc	Positive WHsAg	CC	Rec	Anti-WHc only
A	21	2	2	1	1	1	0
B	25	3	3	2	2	0	1
C	24	10	10	3	1	7	2
D	24	3	3	0	0	3	0
	Totals	18	18	6	4	11	3
	Adults	5	5	3	3	1	1
	Neonates	13	13	3	1	10	2

CC, chronic carriers (WHsAg positive); Rec, recovered (anti-WHs positive); A and B, adult woodchucks; C and D, neonate woodchucks.

years posttransfection. WHsAg and antibodies to WHV core antigen (anti-WHc) were assayed by RIA as described (9, 10). Antibodies to WHsAg (anti-WHs) were detected by a solid-phase RIA similar to that described by Wong *et al.* (11).

RESULTS AND DISCUSSION

Nucleotide Sequence. The complete nucleotide sequence of the WHV8 genome was determined and found to be 3323 nucleotides long (Fig. 2). This value is identical to the previously published genome sizes of the WHV1 (1), WHV7, and WHV59 (4) genomes. In addition to the four known gene sequences (core, polymerase, surface, and X) the WHV8 genome possesses two additional open reading frames (ORFs) recently found in other hepadnavirus genomes—namely, ORF-5 and -6 (12, 13). ORF-5 overlaps the carboxyl terminus of the virus polymerase gene as well as the amino terminus of the virus X gene. ORF-6, on the other hand, overlaps the X and polymerase genes but is located on the opposite, or complementary, DNA strand. Thus, the WHV8 genome resembles that of other previously characterized mammalian hepadnaviruses.

Sequence analysis demonstrates that the genome of WHV8 is most similar to that of the WHV2 genome. The close similarity between WHV2 and WHV8 was expected because both clones were prepared from virions isolated from the serum of the same woodchuck (i.e., WC8). The WHV8 genome differs at 13 (0.4%) nucleotide positions compared to the WHV2 genome. Interestingly, a 3-nucleotide deletion in the polymerase gene sequence previously noted in the sequence of WHV2 (3), but not found in the sequences of other WHV isolates (1, 4), was not present in the WHV8 genome. A comparison between the predicted amino acid sequences of WHV2 and WHV8 demonstrates differences at 19 (1.1%) of the residues of the four established virus genes. In hepadnaviruses, amino acid differences can outnumber nucleotide differences due to the presence of overlapping gene sequences. Overall, however, the nucleotide and predicted amino acid sequences of WHV8 were very similar to those of WHV2, suggesting that the differences between these genomes represent the natural variation among progeny molecules derived from a single parental genome, as previously noted in HBV genomes from infected patients (14, 15).

The nucleotide and predicted amino acid sequences of WHV8 were also very similar to those of WHV7. Both virus isolates were obtained from woodchucks trapped at the same time in the same geographical region (i.e., Maryland). The WHV8 and WHV7 genomes differ at only 18 (0.5%) nucleotide positions. On the other hand, the WHV8 genome differs from the WHV59 and WHV1 genomes at 50 (1.5%) and 102 (3.1%) nucleotide positions respectively. Thus, the extent of nucleotide variation among five WHV genomes is 0.4–3.1%.

Infectivity Assay. Recombinant WHV8 DNA was tested for infectivity in liver transfection experiments using both neonate and adult woodchucks. In four separate experiments, a total of 13 neonate and 5 adult woodchucks were transfected with recombinant WHV8 DNA (Table 1). Evidence of WHV infection was observed in 100% of the transfected animals. anti-WHc appeared in all 18 animals between 2 and 4 months posttransfection. One-third (6/18) of the transfected woodchucks became WHsAg positive, usually coinciding with the first appearance of anti-WHc (Fig. 3). Many of the neonate woodchucks recovered from infection and developed anti-WHs. Of the 6 animals that developed WHsAg (3 neonates and 3 adults), 4 animals (1 neonate and 3 adults) became chronic surface antigen carriers. At 16 months posttransfection, at least 1 animal (WC1806) still possessed $>10^9$ WHV genomes per ml as determined by slot blot hybridization analysis of serum DNA. The above results indicate that the serological patterns following transfection are similar to those observed following experimental inoculation with virus, including the development of chronicity in 22% of the animals (16).

Several groups have suggested that transfected DNA may develop mutations resulting from the action of cellular nucleases (17–20). Therefore, we were interested in deter-

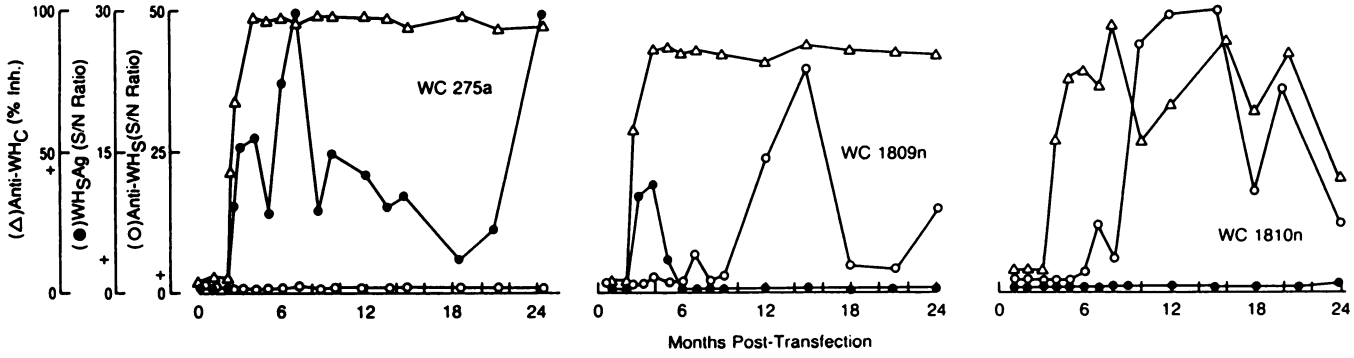

FIG. 3. Serological profile of woodchucks transfected with recombinant WHV8 DNA. Shown are representative serological profiles of an adult (WC275a) and two neonate (WC1809n and WC1810n) woodchucks transfected intrahepatically with monomeric, closed circular, recombinant WHV8 DNA. Anti-WHc is given as percent inhibition (% Inh.), while both WHsAg and anti-WHs are given as a signal/noise (S/N) ratio. Values higher than the "+" on the ordinate axis are considered positive in the assays.

mining whether the transfection procedure would result in a high rate of mutation of the input DNA. Apparently, this was not the case. WHV DNA from serum particles from the chronically infected neonate (WC1806) was cloned and the nucleotide sequence of three independent recombinants was compared directly with that of the input recombinant DNA. The consensus sequence of the three progeny genomes was identical to that of the input parental DNA sequence (R.G. and R.H.M., unpublished data). Therefore, transfection of woodchuck livers with recombinant WHV8 DNA induces active virus replication and gene expression and yields progeny genomes that are faithful copies of the input virus genome.

In conclusion, we have determined the complete nucleotide sequence of the WHV8 genome and demonstrated that this genome is infectious when transfected directly into the livers of susceptible woodchucks. This recombinant should prove useful in attempts to understand hepadnavirus replication and gene expression.

The authors thank B. Baldwin, G. Phillips, and M. Shapiro for assistance with the woodchuck transfections and serology. This work was supported in part under contract numbers NO1-AI-72623 and NO1-AI-82698 from the National Institute of Allergy and Infectious Diseases to Georgetown and Cornell Universities respectively. Computer-assisted nucleic acid and amino acid analysis was through the BIONET National Computer Resource for Molecular Biology, which is supported by National Institutes of Health Grant U41-01685-05.

1. Galibert, F., Chen, T. N. & Mandart, E. (1982) *J. Virol.* **41,** 51–65.
2. Popper, H., Roth, L., Purcell, R. H. & Gerin, J. L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 866–870.
3. Kodama, K., Ogasawara, N., Yoshikawa, H. & Murakami, S. (1985) *J. Virol.* **56,** 978–986.
4. Cohen, J. I., Miller, R. H., Rosenblum, B., Denniston, K., Gerin, J. L. & Purcell, R. H. (1988) *Virology* **162,** 12–20.
5. Denniston, K., Gerin, J. L. & Purcell, R. H. (1988) *Virology* **162,** 12–20.
6. Dubensky, T. W., Campbell, B. A. & Villarreal, L. P. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7529–7533.
7. Seeger, C., Ganem, D. & Varmus, H. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5849–5852.
8. Zagursky, R. J., Baumeister, K., Lomax, N. & Berman, M. L. (1985) *Gene Anal. Tech.* **2,** 89–94.
9. Cote, P. J., Dapoloto, G. M., Shih, J. & Gerin, J. L. (1984) *J. Virol.* **49,** 701–708.
10. Ponzetto, A., Cote, P. J., Ford, E. C., Engle, R., Cicmanec, D., Shapiro, M., Purcell, R. H. & Gerin, J. L. (1985) *Virus Res.* **2,** 301–315.
11. Wong, D. C., Shih, J., Purcell, R. H. & Gerin, J. L. (1982) *J. Clin. Microbiol.* **15,** 484–490.
12. Miller, R. H. (1988) *Science* **239,** 1420–1422.
13. Kaneko, S. & Miller, R. H. (1988) *J. Virol.* **62,** 3979–3984.
14. Okamoto, H., Imai, M., Kametani, M., Nakamura, T. & Mayumi, M. (1987) *Jpn. J. Exp. Med.* **57,** 231–236.
15. Okamoto, H., Tsuda, F. & Mayumi, M. (1987) *Jpn. J. Exp. Med.* **57,** 217–221.
16. Ponzetto, A., Cote, P. J., Ford, E. C., Purcell, R. H. & Gerin, J. L. (1984) *J. Virol.* **52,** 70–76.
17. Hauser, J., Levine, A. S. & Dixon, K. (1987) *EMBO J.* **6,** 63–67.
18. Miller, J. H., Lebkowski, J. S., Greisen, K. S. & Calos, M. P. (1984) *EMBO J.* **3,** 3117–3121.
19. Razzaque, A., Chakrabarti, S., Joffee, S. & Seidman, M. (1984) *Mol. Cell. Biol.* **4,** 435–441.
20. Wake, C. T., Gudewicz, T., Porter, T., White, A. & Wilson, J. H. (1984) *Mol. Cell. Biol.* **4,** 387–398.