ISSN 1009 - 3095 *Journal of Zhejiang University( SCIENCE)* V.1, No.2, P.207 - 211, Apr. - Jun., 2000
http://www.zju.edu.cn; http://www.zjupress.com; http://www.chinainfo.gov.cn/periodical 

# MOLECULAR CLONING AND NUCLEOTIDES SEQUENCE ANALYSIS OF $G_1$ GENOME SEGMENT OF HANTAAN VIRUS $Z_{10}$ STRAIN

LIANG Wei-feng(梁伟峰)[1], SHEN Yue-hong(沈月洪)[1],
MA Yi-lin(马亦林)[1], ZHAO Nian-feng (赵年丰)[1], CHEN Ya-gang(陈亚岗)[1],
ZHU Zhi-yong (朱智勇)[2], XU Xiao(徐晓)[3]

([1] *The First Affiliated Hospital of Zhejiang University, Hangzhou, 310003, China*)
([2] *Zhejiang Health and Anti-epidemic Station, Hangzhou 310006 China*)
([3] *University of California, USA*)

Received May.20, 1999; revision accepted Aug.15, 1999

**Abstract:** The molecular cloning of the $G_1$ genome segment of the $Z_{10}$ strain of Hantaan virus and analysis of the first gene data on the currently widely used Hantavirus vaccine strain in China are presented. The $G_1$ genome segment of $Z_{10}$ virus was amplified by the method of RT-PCR, and the products were cloned into the pGEM-T vector after identification and purification. The clone was sequenced by Sanger's dideoxy chain termination method. The 1449 bases of the $Z_{10}$ $G_1$ genome segment, and the coding for 483 amino acids were determined. The base compositions for the $G_1$ segment RNA determined from cDNA sequence information, were 20.6% A, 21.6% G, 18.8% C and 30.0% U. These values are similar to those of the 76/118, Lee and SR virus. As compared to the $Z_{10}$ $G_1$ segment, the sequence homology at the nucleotide level is 87% (76/118, type Ⅰ), 86% (Lee, type Ⅰ), 86% (Hojo, type Ⅰ), 67% (R22, type Ⅱ), and 59% (K22, type Ⅲ). The $Z_{10}$ virus $G_1$ protein amino acid sequence identity with other Hantaan viruses (94 - 95% homology) was higher than that with Seoul type viruses (77 - 80%). More amino acid sequence heterogeneity between the $Z_{10}$ and 76/118 was observed in the N terminal, especially the diverging cluster of ten amino acids at position 84 to 93 of the $Z_{10}$ virus $G_1$ protein. Conclusion: (1) The $Z_{10}$ strain is one of the Hantaan viruses. (2) The important region of the $Z_{10}$ $G_1$ segment was conservative. (3) Although substantial divergence of nucleotide sequences of the $G_1$ genome segment was found between the $Z_{10}$ strain and other type I Hantaviruses, relatively high amino acid sequence homology was shown among them. Thus, good immune protection could be obtained with the inactivated Meriones unguiculatus kidney cell vaccine against other strains of Hantaviruses.

**Key words:** Hantaan virus, reverse transcription-PCR, sequence analysis
**Document code:** A **CLC number:** R373.3+2

## INTRODUCTION

Hemorrhagic fever with renal syndrome (HFRS) is a natural focal disease caused by viruses in the Hantavirus genus of the family Bunyaviridae. At least eight serologically distinct Hantaviruses were recognized. HFRS is endemic in China and is only caused by Hantaan Virus (type Ⅰ) transmitted by *Apodemus agrarius* and Seoul virus (type Ⅱ) by *Rattus norvegicus*. Hantavirus has a single-stranded, three-segmented, negative-sense RNA genome. The L segment is suggested to encode the viral polymerase. It was shown that the S segment encodes the nucleocapsid protein and that the M segment encodes the two glycoproteins $G_1$ and $G_2$.

$Z_{10}$ strain was isolated from a patient with HFRS in Sheng City, Zhejiang Province, China. The virus (Zhu, 1991) is characterized by high titer of haemagglutinin, high level of virus titer and a long period in cell cultures. Mongolian Gerbils Kidney tissue culture inactivated HFRS vaccine, $Z_{10}$ strain as a seed virus, is safe、effective and stable, and approved by the Chinese Ministry of Health. $Z_{10}$ virus is considered as an ideal seed virus for HFRS inactivated vaccine. The gene products of the Hantaan virus M segment, especially the $G_1$ segment, have many of the important biological properties of the virus (Euiott, 1990; Shope et al., 1981), including virulence, neutralization, haemagglutination and cell fusion. In this paper, we present the molecular cloning and analysis on the first gene data of the currently widely used Hantavirus vaccine

strain in China.

## MATERIALS AND METHODS

**Virus and cell** Hantaan virus $Z_{10}$ strain was propagated in Vero E6 cells (C1008, ATCC), which were grown in Eagle's minimum essential medium (EME). Virus particles were pelleted by ultracentrifugation.

**RNA extraction** RNA was extracted by TRIzol (Promega Co., USA) as described in the guideline.

**cDNA synthesis** First-strand cDNA synthesis was primed with $P_1$ primer (5'ACCGGATCCATAGTAGTAGACTCCGCA 3' 1nt ~ 27nt) using AMV reverse transcriptase (Life Sciences Inc., USA). The 20 μl solution contained 1 μg of sample RNA, 2 μl of 10(reverse transcriptase buffer, 4 μl of 2 mmol/L $MgCl_2$, 2 μl of 10 mmol/L dNTP, 0.5 μg of the primer, and 15 units of AMV reverse transcriptase. The mixture was incubated at 42 ℃ for 1 hour.

**PCR amplification** An RNA-DNA hybrid was used for PCR amplification using primer $P_1$ and $P_2$ (5' AATAGAATGTGCTACTCCCAG 3' 1461nt ~ 1480nt). The PCR product was identified by nested PCR using primer $P_3$ (5' CAATCAGCAACATGGGGATA 3' 30nt ~ 49nt) and P4 (5' AATATCAAAGATCCCATG 3' 631nt ~ 648nt). The 50 μl PCR solution contained 5 μl of 10× buffer, 3 μl of 25 mmol/L $MgCl_2$, 5 μl of 10 mmol/L dNTP, 2 μl of 10 pmol/L primer, 4 μl of RT-product, and 2 units of Taq DNA polymerase (Promega Co., USA) PCR was done through 35 cycles (94 ℃ for 1 minute, 55 ℃ for 1 minute, and 72 ℃ for 1.5 minutes).

**Molecular cloning** After the PCR products were purified by 1% low-melting agarose gel, they were ligated into pGEM-T vectors following the manufacture's instructions (Promega Co., USA). The subcloning and sequencing strategy is shown in Fig.1.

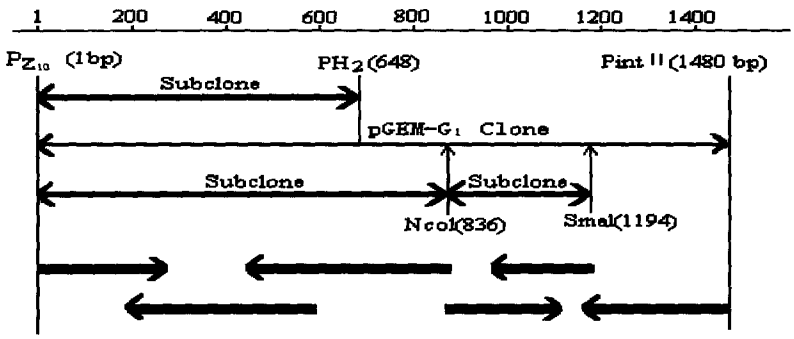


**Fig.1 Subcloning and sequencing strategy of $Z_{10}$ virus $G_1$ segment cDNA clone. The arrows represent the direction of sequencing.**

**Nucleotide sequence analysis** The clone was sequenced by Sanger's dideoxy chain termination method with a sequenase kit (United States Biochemical Co.). Nucleotide and amino acid sequence were compiled and analyzed by the Genbank and PDB program.

## RESULTS

**Nucleotide sequence analysis** Fig.2 shows 1449 bases of $Z_{10}$ $G_1$ genome segment bases, which can code 483 amino acid. The base composition of the $G_1$ segment RNA was determined from cDNA sequence information to be 20.6% A, 21.6% G, 18.8% C and 30.0% U. These values are similar to those of the 76/118, Lee and SR virus. The $Z_{10}$ $G_1$ segment's sequence homology at the nucleotide level was 87% (76/118, type Ⅰ), 86% (Lee, type Ⅰ), 86% (Hojo, type Ⅰ), 67% (R22, type Ⅱ), and 59% (K22, type Ⅲ).

**Coding strategy of the $Z_{10}$ $G_1$ RNA segment (ORF)** A single long open reading frame (ORF) was detected in the nucleotide sequence corresponding to viral complementary-sense RNA. Within the major ORF, protein synthesis could potentially begin at either of two in-frame initiation codons at nucleotide position 40 – 42 or 64 – 66. According to Kozak (1978),

```
      |  10       |  20       |  30       |  40       |  50       |  60
    1 TAGTAGTAGA TCCGCAAAAG AAAGCAGTCA ATCAGCAACA TGGGGATATG GAAGTGGCTA  60
   61 GTAATGGCCA GTTTAGCATG GCCAGCTTAC ACACTACGAA ATGTGTATGA CATGAAAATT 120
  121 GAGTGCCCTC ACACAGTCAG CTTTGGTGAA AACAGCGTGA TAGGTTATGT GGAATTGCCC 180
  181 CCTATGCCAT TGGCTGATAC AGCACAACTG GTGCCTGAAA GTTCTTGTAA CATGGACAAC 240
  241 CACCAATCAA TGAACACAAT AACAAAATAC ACCCAGTTAA GTTGGAGGGT AAGGCTGAAC 300
  301 AGGCACAAGC CAGCCAAAAA TTCATTTGAG GCAGTGTCCA CTGAAGTTTA CTTGAAAGGA 360
  361 ACTTGTGTTT TGAAGCACAA AATGCTAGAG GAGTCATACC GCAGTAGGAA ATCAATAACT 420
  421 TGTTATGACC TTTCTTGTAA CAGCACTTAC TGTAAGCCCA CCCTGTACAT GATTGTACCA 480
  481 ATTCATGCAT GTAACATGAT GAAAAGTTGT TTGATTGCAC TGGGCTTACA CAGAGTACAA 540
  541 GTGGTGTATG AAAGAACCTA CTGCATGACA GGGGTCCTGA TAGAAGGAAA ATGTTTTGTC 600
  601 CCAGATCAAA GTGTGGTCAG TATTATCAAG CATGGGATCT TTGATATTGC AAGTGTTCAT 660
  661 ATTGTATGTT TCTTTGTTGC AGTTAAAGGG AATACTTATA AAATTTTTGA ACAGGTTAAG 720
  721 AAATCCTTTG AATCAACATG CAATGATACA GATAATAAAG TGCAAGGATA TTACATTTGC 780
  781 ATTGTGGGGG GAAACTCTGC ACCAATATAT GTACCAACAC TTGGTGATTT CAGATCCATG 840
  841 GAGGCATTTA CAGGAATCTT TAGGTCACCA CATGGGGAGG GCCATGATCT AGCTGGGGAA 900
  901 GAGATTGCAT CTTACTCTAT TGTTGGACCT GCCAATGCAA AAGTGCCTCA CAGAGCTAGC 960
  961 TCAGACACAT TAAGCTGGAA TGCCTATTCA GGTATACCAT CCTATTCTTC ACTCAGCATC 1020
 1021 CTTGCAGGCT CAACAGAAGC TAAGCATGTA TTTAGCCCTG GTTTTTTTCC ACAGCTTAAT 1080
 1081 CACACGAAGT GTGACAAGAC TGCTATCCCA CTCACTTGGA CAGGGATGAT TGACTTGCCT 1140
 1141 GGGTACTATG AGGCTATACA TCCTTGTACA GTTTTCTGTG TTTTATCAGG CCCCGGGGCA 1200
 1201 TCCTGTGAAG CATTCTCTGA AGGTGGGATC TTCAACATAA CATCCCCTAT GTGCTTAGTA 1260
 1261 TCAAAACAAA ATCGATTCCG GTTGACAGAG CAGCAAGTAA ACTTTGTCTG TCAAAGGGTT 1320
 1321 GATATGGATA TTATTGTGTA TTGTAATGGC CAGAGGAA. AG TGATATTAAC AAAAACTTTA 1380
 1381 GTCATAGGAC AGTGTATATA TACTATAACA AGTTTATTTT CACTACTGCC TGGAGTAGCA 1440
 1441CATTCTATT                                                          1449
      |  10       |  20       |  30       |  40       |  50       |  60
```

**Fig.2 Nucleotide sequence of the $Z_{10}$ virus $G_1$ RNA segment, shown as the viral complementary DNA(5' to 3'). The initiation codon (ATG) is underlined.**

the first codon has more favorable flanking sequences for initiation of protein synthesis. Examination of all six possible reading frames of the cDNA revealed additional ORFs initiating ATG cordons and encoding seven peptides, more than 100 amino acids, the longest was 306 amino acids, but there was no frame beyond 100 amino acids in 76/118 strain (Schmal et al., 1987).

**GS Site** Table 1 presents results of a computer aided search for potential asparagine-linked glycosylation sites (Asn-X-Ser/Thr, GS). The four GS sites in the $Z_{10}$ virus were conserved in the other type I virus (76/118, Lee, Hojo, HV114). Two of the four GS positions were conserved in type II virus (R22 B1, HR 80 - 39). No GS site was conservative in type III (Yoo et al., 1987).

**Amino sequence analysis** The deduced amino sequences of the $G_1$ protein are compared in Fig.3. The 17 amino acids following the first AUG of the Hantaan M segment ORF and preceeding the amino terminus of the mature $G_1$, constitute a typical glycoprotein signal sequence. The $Z_{10}$ virus $G_1$ protein had higher amino acid sequence identity with other Hantaan viruses (94 - 95% homology) than with Seoul type viruses (77 - 80%).

More amino acid sequence heterogeneity between the $Z_{10}$ and 76/118 was observed in the N terminal. The biggest variable region was identi-

fied at the amino acid positions 84 to 93 in the $G_1$ protein of $Z_{10}$ and 76/118 virus.

**Table 1 Comparison of homology of the nucleotide sequence, deduced amino sequence and potential asparagin-linked glycosylation sites of the $G_1$ glycoprotein encoding region of several typical strains.**

| Serotype | I (HTNY) | | | | II (SEOV) | | | III (PUUV) | | IV (PHV) | V (TV) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Strain | 76/118 | Lee | Hojo | HV114 | HR80-39 | B-1 | R22 | Sotkamo | K27 | PH-1 | Thai749 |
| Nucleotide(%) | 87 | 86 | 86 | ND* | 68 | ND | 67 | 60 | 59 | ND | 67 |
| Amino(%) | 95 | 94 | 94 | 95 | 80 | 80 | 77 | 71 | 70 | 73 | 82 |
| GS site | 4 | 4 | 4 | 4 | 2 | 2 | 2 | 0 | 0 | 1 | 2 |

* ND: no data

Analysis of potential transmembrane helices identified only one conserved region in the $G_1$ protein of Hantavirus (Mohana et al., 1986). A completely conservative region was recognized between Hantaan virus (76/118 Lee. Hojo HV114) and Seoul virus (B1, HR80-39) except R22 strain.

The cleavage between the signal peptide and the N terminus of $G_1$ was determined to be at 17-18 site in $Z_{10}$, at amino acid position 18-19 in 76/118.

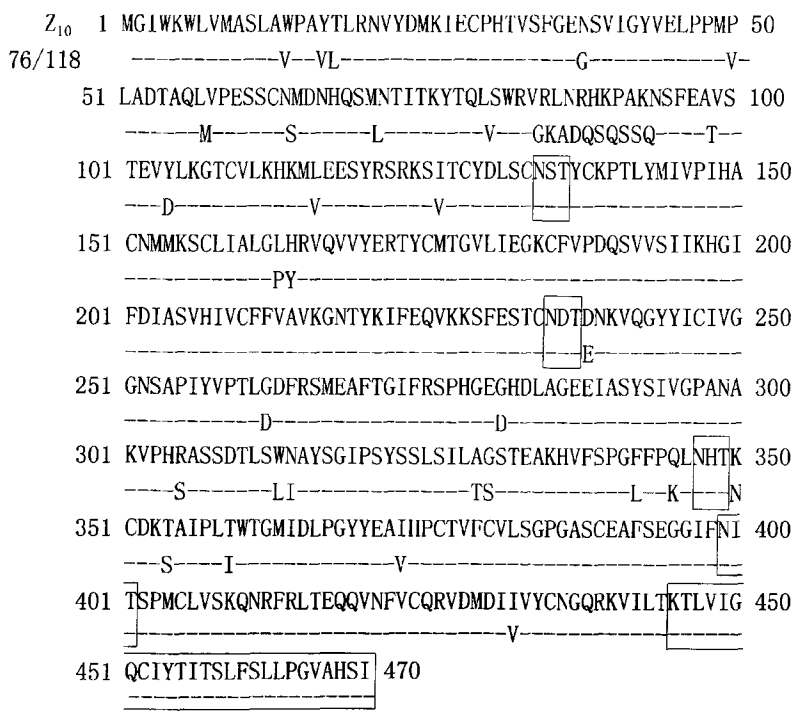


**Fig. 3 Amino acid sequence comparison of $Z_{10}$ $G_1$ and 76/118 $G_1$ sequence-encoded polypeptides. $Z_{10}$ $G_1$ encoded polypeptides are presented above, 76/118 below. Identical amino acids are denoted by bars. The small boxes denote potential asparagine-linked glycosylation sites. The large box denotes the potential transmembrane domain**

## DISCUSSION

Nucleotide sequence analysis of the $G_1$ genome segment of Hantaan virus $Z_{10}$ strain revealed that $Z_{10}$ had the common Hantaan virus property. Hantaan virus $Z_{10}$ strain also apparently had the common (Bunyaviridae) property of a leader sequence before the encoded glycoprotein, as had been observed for the phlebo-and bunyaviruses (Schmaljohn et al., 1985). The

important region of the $G_1$ segment of $Z_{10}$ was conservative.

Various Hantaviruses were isolated from HFRS patients and various rodent species in many parts of the world. The nucleotide substitution frequencies for 76/118 and Seoul quasispecies were $1.7 \times 10^{-4}$ and $3.3 \times 10^{-4}$ respectively (Arikawa et al., 1990).

As compared with 76/118, Lee and R22 virus, the large number of amino acids, especially the divergence of ten amino acids at position 84 to 93 in the $Z_{10}$ virus $G_1$ protein, may contribute much to the high level of the viral replication and haemagglutinin, the stability of the virus structure, and glycoprotein, and better immunogenicity; but may also alter the viral infections ability, toxicity and genetic stability (Hopp et al., 1981).

Mechanisms of Hantavirus evolution (Newton et al., 1981) (accumulation of point mutations, deletions/insertions, reassortment of genome RNA segments, putative intracistronic recombination and quasi-species) are being studied.

To our knowledge, there is no report of person to person transmission of Hantavirus and no report of chronic patients in literature. The hantavirus-human interaction is very short. So it is impossible the mutation is caused by human immunity pressure (Lee et al., 1978).

Coevolution of Hantavirus and their reservoirs is now being paid more attention. Hantaviruses are rodent-borne human pathogens. Each of the well-characterized viruses in the genus is associated with a different primary rodent reservoir. The worldwide appearance of hemorrhage fever with renal syndrome (HFRS) mainly follows the distribution and population dynamics of the rodents carrying the virus. Genetic diversity was shown to correlate more to the geography of mice trapping rather than to the year of isolation. This divergence might have been caused by adaptation to the host (coevolution) since rodent species, not geography, determine the stereotypic profile (Bilsel et al., 1988).

The HFRS vaccine efficacy is evaluated with Hantaan virus 76/118, which as a standard virus, attacks the animal that was immunized with the vaccine in China. Although substantial divergence of nucleotide sequences of the $G_1$ genome segment was found between the $Z_{10}$ strain and other type I Hantaviruses, such as 76/118 strain, relatively high amino acid sequence homology was shown among them. Thus, good immune protection could be obtained with the inactivated *Meriones unguiculatus* kidney cell vaccine against other strains of Hantaviruses.

## References


Arikawa, J., Lapenotiere, H. F., Iacono-Connors, L., et al., 1990. Coding properties of the agents of hemorrhagic fever with renal syndrome. *Virology*, **176**:114 - 125.

Bilsel, P. A., Tesh, R. B., Nichol, S. T., 1988. RNA genome stability of Toscana virus during serial transovarial transmission in the sandfly *Phlebotomus perniciosus*. *Virus Res*, **11**:87 - 94.

Elliott, R. M., 1990. Molecular biology of the Bunyaviridae. *J Gen Virol*, **71**: 501 - 522.

Hopp, T. P., Woods, K. R., 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA*, **78**:3824 - 3828.

Kozak, M., 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell*, **15**:1109 - 1123.

Lee, H. W., Lee, P. W., Johnson, K. M., 1978. Isolation of the etiologic agent of Korean Hemorrhagic fever. *J Infect Dis*, **137**:298 - 308.

Mohana Rao, J. K., Argos, P., 1986. A conformational preference parameter to predict helices in integral membrane proteins. *Biochim Biophys Acta*, **869**:197 - 214.

Newton, S. E., Short, N. J., Dalgarno, L., 1981. Bunyaviridae virus replication in cultured *Aedes albopictus* (mosquito) cells: establishment of a persistent viral infection. *J Virol*, **38**:1015 - 1024.

Schmaljohn, C. S., Hasty, S. E., Dalrymple, J. M., et al., 1985. Antigenic and genetic properties of viruses linked to hemorrhagic fever with renal syndrome. *Science*, **227**:1041 - 1044.

Schmaljohn, C. S., Schmaljohn, A. L., Dalrymple, J. M., 1987. Hantaan virus M RNA: coding strategy, nucleotide sequence, and gene order. *Virology*, **157**:31 - 39.

Shope, R. E., Rozhon, E. J., Bishop, D. H., 1981. Role of the middle-sized bunyavirus RNA segment in mouse virulence. *Virology*, **114**:273 - 276.

Yoo, D. W., Kang, C. Y., 1987. Nucleotide sequence of the M segment of the genomic RNA of Hantaan virus 76 - 118. *Nucleic Acids Res*, **15**:6299 - 6300.

Zhu, Z. Y., Tong, H. Y., Li, Y. J., et al., 1991. The seed choice of inactive epidemic hemorrhage fever. *The journal of Chinese Microbiology and Immunology*, **11**:155 - 157. (In Chinese, with English abstract)