

# Variations in genome fragments coding for RNA polymerase in human and simian hepatitis A viruses

M.S. Balayan, Yu.Yu. Kusov, A.G. Andjaparidze, S.A. Tsarev, E.D. Sverdlov, V.E. Chizhikov, V.M. Blinov and S.K. Vasilenko

*Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, M.M. Shemyakin Institute of Bioorganic Chemistry, Moscow and Institute of Molecular Biology, Novosibirsk, USSR*

Received 24 February 1989

The genome of hepatitis A virus (HAV) isolated from spontaneously infected African vervet monkey (*Cercopithecus aethiops*) has been cloned and partially sequenced. Comparison of genome fragments (1248 and 162 bp) from the 3D (RNA polymerase) region with the corresponding parts of human HAV genomes revealed a high degree of heterogeneity: there were altogether 257 nucleotide changes leading to 44 substitutions in predicted amino acid sequence, i.e. 89% amino acid identity. This divergence is considered to be significantly greater than genomic variations usually found among human HAV strains, where amino acid identity in the 3D region is over 98%.

Hepatitis A virus; Genomic variation; Nucleotide sequence; Picornavirus; RNA polymerase

## 1. INTRODUCTION

In recent years several hepatitis A virus (HAV) strains originally isolated from humans have been cloned and completely or partially sequenced [1-7]. Sequence analysis has shown a high degree of genomic homology among strains of diverse geographic origin as well as between the 'wild' and cell-culture-adapted variants [6-8]. The HAV strains so far compared appeared to have identity in greater than 90% of nucleotides and almost 98% of predicted amino acids [9]. Antigenic comparison

has also demonstrated a close relatedness of all known HAV strains (review [10]).

It has become evident that spontaneous hepatitis A infection occurs in certain species of non-human primates including New [11] and Old [12] World monkeys. The virus isolated from infected animals shared the main antigenic characteristics of human HAV, but could be differentiated with some monoclonal antibodies [13].

We report here a first attempt at cloning the genome of HAV isolated from a spontaneously infected African vervet monkey. Genome fragments available for sequence analysis corresponded to the cDNA part coding for RNA polymerase (3D protein).

*Correspondence address:* M.S. Balayan, Institute of Poliomyelitis and Viral Encephalitides, 142782 Moscow, USSR

*Abbreviations:* PV1, poliovirus type 1; CVB4, Coxsackie virus B4; HRV14, human rhinovirus type 14; HRV2, human rhinovirus type 2; EMCV, encephalomyocarditis virus; FMDV, foot-and-mouth disease virus; huHAV, human hepatitis A virus; siHAV, simian hepatitis A virus

The nucleotide sequences presented here have been submitted to the EMBL/GenBank database under the accession no. Y07509

## 2. MATERIALS AND METHODS

### 2.1. Virus

Spontaneously infected African vervet monkey (*Cercopithecus aethiops*) was killed at the peak of the disease (serum alanine aminotransferase elevation). The presence of HAV particles in the liver has been demonstrated by immuno-electron microscopy, immunofluorescence, and enzyme immunoassay. A 15% (w/v) suspension of liver tissue (total, 750 g) was prepared and

treated with liquid N<sub>2</sub> and freon-113 followed by low-speed centrifugation. Virus liberated into the aqueous phase was pelleted on a CsCl cushion (1.40 g/cm<sup>3</sup>) and further purified by consecutive isopycnic and rate-zonal centrifugation in CsCl and sucrose gradients, respectively. Finally, virus was concentrated by pelleting on the bottom of centrifuge tubes and resuspending in minimal volume of 0.1 M PBS. During the entire procedure the virus was monitored by enzyme immunoassay and immunoelectron microscopy.

RNA was extracted from virus particles using hot phenol and chloroform after pre-incubation with proteinase K (500 µg/ml) in the presence of SDS (0.5%). The amount of viral RNA in the extract was calculated to be 10 ng.

## 2.2. Cloning

Viral RNA was reverse-transcribed into single-stranded cDNA using oligo(dT)<sub>12-18</sub> as primer. The second strand was obtained by *E. coli* DNA polymerase I directed synthesis in the presence of RNase H, yielding approx. 0.2 µg double-stranded cDNA. *Eco*RI linkers were tailed to this cDNA and then cleaved with *Eco*RI.

To remove excess linkers cDNA was mixed with tRNA at final concentration of 0.5 mg/ml and twice precipitated with an equal volume of isopropanol followed by 1 h incubation and centrifugation. The cDNA (2 ng) was inserted into λgt 10 phage

DNA cleaved with *Eco*RI (overnight at +4°C). The concatemers obtained were packaged in vitro into λ phage proteins, reaching an efficiency of packaging of 2–5 × 10<sup>7</sup> pfu/µg insertion. The selective BNN 102 strain of *E. coli* [14] was used for plating with vector background no greater than 5%. To generate the clone library 2 × 10<sup>5</sup> individual clones were plated at a density of 5 × 10<sup>4</sup> colonies per 100-mm plate. Amplification was achieved by elution of phages into SM medium for 16 h. The amplified library available for screening has a titre of 10<sup>11</sup> pfu/ml. Phage recombinants containing the HAV sequences were identified via DNA-DNA colony-blot hybridization with nick-translated <sup>32</sup>P-labelled DNAs complementary to three different parts of the human HAV genome from a plasmid library described elsewhere [15]. For determination of the primary structure of inserts they were re-cloned into the *Eco*RI site of pUC19 plasmid

## 2.3. Sequencing and sequence analysis

Nucleotide sequences of the inserts were determined using the procedure of Maxam and Gilbert [16]. Amino acid sequences of picornavirus RNA polymerases were aligned by OPTAL programme [17] run on an ES-1060 computer. Putative enzyme active sites for HAV RNA polymerase were identified from alignment of amino acid sequences of HAV with those of other picornaviruses.

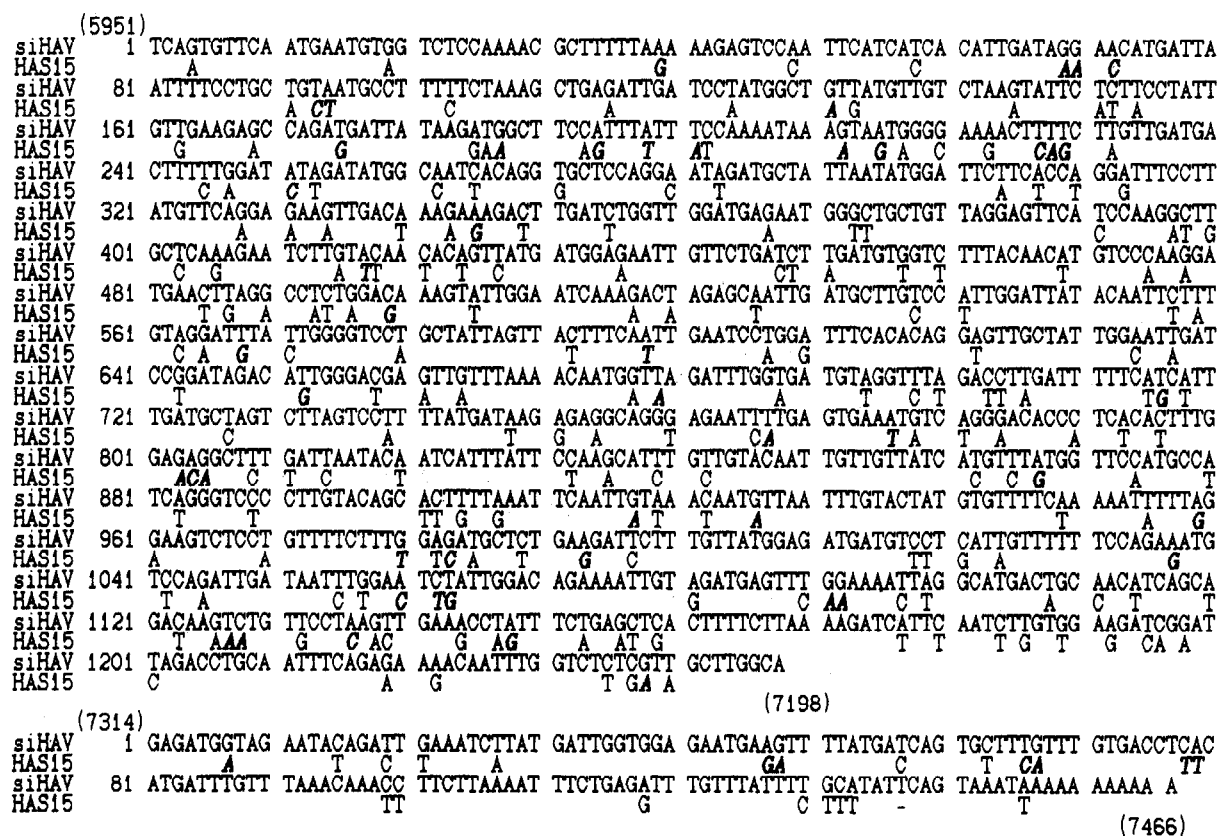


Fig. 1. Nucleotide sequences of cDNA fragments from the 3D (RNA polymerase) region. Comparison of simian (siHAV) and human (HAS15) HAVs (italicized letters denote changes resulting in differences in predicted amino acid sequence).

## 3. RESULTS

Two cDNA fragments identified in phage recombinants contained respectively 1248 and 162 bp representing the region of the genome encoding 3D protein, i.e. RNA polymerase. According to the numbering system suggested by Najarian et al. [4] for the human HAV genome, one fragment is positioned between nucleotides 5951 and 7198 inclusive, the other from nucleotide 7314 onwards. The latter carries the sequence corresponding to 144 nucleotides at the 3'-end, including the non-translated region and part of the poly(A) tract.

Both fragments cover almost 95% of the 3D region.

The nucleotide composition of these fragments differs from the corresponding parts of the human HAV genome, HAS-15 strain [18], by altogether 257 scattered changes (81.5% identity) which result in 44 differences in the predicted amino acid sequence (fig.1).

Computer-aided comparison of the predicted amino acid sequences with those of other picornaviruses (fig.2) revealed that, despite the high degree of genomic heterogeneity in this region, the substitutions found in simian HAV RNA apparently

|   |       |  |  |
|---|-------|--|--|
| 1 | PV1   | GCDPDLEWSKIFVLME--EK-LEAEDYTGSDASLSPAWEAL--KMLEKIGFGDRVD--   |  |
| 2 | CVB4  | GCDPDCFWSKIPVMLD--GH-CIAFDYSGYDASLSPVWFACL--KMLCEKCGYTHKETN- |  |
| 3 | HRV14 | GCDPDVFWSVIPCLMD--GH-LMAFDYSNFDASLSPVWFVCL--EKVLTCLGFAGSS--- |  |
| 4 | HRV2  | GCDPEVFWSKIPAMLD--DKCIMAEDYTNVDGSIHPWFEL--KGVLDLDFNPT----    |  |
| 5 | EMCV  | GCDPDVHWTAFGVAMQGF-ERVYDVDYSNFDSTHSVAMFRLLAEEFFTPENGFDPLTRE- |  |
| 6 | FMDV  | GCNPDVDWQRFQTHFAQY-RNVWDVDYSAFDANHCSAMNIMFEEVFRDFTGFHPNAEW-  |  |
| 7 | huHAV | GIDPDRQWDELFKTMIRFGDVGLDLDFSAFDASLSPFMIREAG-RIMSELSGTPSHFGTA |  |
| 8 | siHAV | GIDPDRHWDELFKTMVRFQDVGLDLDFSSFDASLSPFMIREAG-RILSEMSGTPSHFGEA |  |
|   |       |  |  |
| 1 | PV1   | YIDYLNHSHHLYKNKTYCVKGGMPSGCSGTSIFNSMINNLIIRTLKTYKG---IDL     |  |
| 2 | CVB4  | YIDYLCNSHHLYRNKHYFVRGGMPSGCSGTSIFNSMINNLIIRTLMLKVYKG---IDL   |  |
| 3 | HRV14 | LIQSICNTHHIFRDEIYVVEGGMPSGCSGTSIFNSMINNLIIRTLILDAYKG---IDL   |  |
| 4 | HRV2  | LIDRLCKSKHIFKNTYYEVEGGVPSGCSGTSIFNTMINNLIIRTLVLDAYKN---IDL   |  |
| 5 | EMCV  | YLESIAISTHAFEEKRFLITGGLPSCAATSMNLNTIMNNLIIRAGLYLTYKN---FEF   |  |
| 6 | FMDV  | ILKTLVNTTEHAYENKRIIVEGGMPSGCSATSINTILNNIYVLYALRRHYEG---VEL   |  |
| 7 | huHAV | LINTIIYSKHLLYNCCYHVCVSGMPSGSPCTALLNSIINNINLYYVFSKIFRKSPVFFG  |  |
| 8 | siHAV | LINTIIYSKHLLYNCCYHVYVSGMPSGSPCTALLNSIVNNVNLYYVFSKIFRKSPVFFG  |  |
|   |       |  |  |
| 1 | PV1   | LKMIAYGDDVIA--SYPHEVD-ASLLAQS---GKDYGLTMTPADKSA-IF-ETVTWEN   |  |
| 2 | CBV4  | FRMIAYGDDVIA--SYPWPID-ASLLAEA---GKGYGLIMTPADKGE-CF-NEVTWTN   |  |
| 3 | HRV14 | LKILAYGDDLIV--SYPYELD-PDVLATL---GKNYGLTITPPDKSE-TF-TKMTWEN   |  |
| 4 | HRV2  | LKIIAYGDDVIF--SYIHELD-MEAI AIE---GVKYGLTITPADKSN-TF-VKLDYSN  |  |
| 5 | EMCV  | VKVL SYGDDLLV--ATNYQLD-FDKVRAS---LAKTGYKITPANTTS-TFPLNSTLED  |  |
| 6 | FMDV  | YIMISYGDDIVV--ASDYDLDFEALKPH---FKSLGQTITPADKSDKGFVLGQSLTD    |  |
| 7 | huHAV | LRILCYGDDVLIVFSRDVQIDNLDLIGQKIVDEFKKGMTATSADKN---VPQLKPVSE   |  |
| 8 | siHAV | LKILCYGDDVLIVFSRNVDNLESIGQKIVDEFGKLGMTATSADKS---VPKLKPISE    |  |
|   |       |  |  |
| 1 | PV1   | VTFLKRRFFRADEKYPFLIHPVMPMKEHESI-RW                           |  |
| 2 | CVB4  | VTFLKRYFRADEQYPFLVHPAMPMDHESI-RW                             |  |
| 3 | HRV14 | LTFLKRYFKRDQQTFLVHPVMPMKDHESI-RW                             |  |
| 4 | HRV2  | VTFLKRGFKQDEKYNFLIHPTFPEDEFESI-RW                            |  |
| 5 | EMCV  | VVFIKRKFKE--GP-LYRPVMNREA-EAMLSY                             |  |
| 6 | FMDV  | VTFLKRHFHMD-YGTGFYKPMASKT-EAILSF                             |  |
| 7 | huHAV | LTFLKRFSNLVE---DRIRPAISEKTIWSLIAW                            |  |
| 8 | siHAV | LTFLKRFSNLVE---DRIRPAISEKTIWSLVAW                            |  |

Fig.2. Amino acid alignment of fragment from the 3D (RNA polymerase) region for simian HAV with other picornaviruses (conserved amino acids forming the putative active site of the enzyme are indicated th the large dots).

do not alter the putative active site of the enzyme. Conserved amino acids presumably responsible for enzymatic function of the 3D protein remain unchanged in the simian HAV.

#### 4. DISCUSSION

The HAVs of simian origin have not yet been intensively studied. In the only previous paper on the same subject, genomic heterogeneity among human and non-human HAV strains has been demonstrated by the failure to hybridize simian HAV RNAs with cDNA probes derived from the human HAV genome [19]. The differences seem to be located in the region coding for virus structural proteins (VP1 and VP2).

The significance of variations in the region coding for non-structural proteins remains obscure. It is of interest that among the human HAV strains compared to date, differences in the 3D (RNA polymerase) region do not exceed 7–9 substitutions in amino acid sequences (over 98% identity) [6–8]. In contrast, the simian HAV examined here exhibits significant differences from human HAV in the amino acid composition of the 3D region: the number of substitutions was found to be 44 in the fragments representing approx. 95% of the length of this region, equalling 90% amino acid identity. A comparison of human HAV with poliovirus type 1 Mahoney strain showed the amino acid identity of the 3D region to be 29% [9]. Despite the poor degree of homology in the regions coding for structural [19] as well as non-structural (this work) proteins, human and simian HAVs retain close antigenic relatedness and presumably similar enzymatic (RNA polymerase) function. This observation prompts us to investigate the evolutionary relationship among HAVs originating from different hosts in the order of Primates.

*Acknowledgements:* The authors are indebted to Dr V.I. Grabko for his assistance in the procedure of virus purification. Technical assistance by Valentina Gorbachkova in preparation of the manuscript is greatly appreciated.

#### REFERENCES

- [1] Ticehurst, J.R., Racaniello, V.R., Baroudy, B.M., Baltimore, D., Purcell, R.H. and Feinstone, S.M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 5885–5889.
- [2] Linemeyer, D.L., Menke, J.G., Martin-Gallardo, A., Hughes, J.V., Young, A. and Mitra, S.W. (1985) *J. Virol.* 54, 247–255.
- [3] Baroudy, B.M., Ticehurst, J.R., Miele, T.A., Maizel, J.V., Purcell, R.H. and Feinstone, S.M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2143–2147.
- [4] Najarian, R., Gaput, D., Gee, W., Potter, S.J., Renard, A., Marryweather, J., Van Nest, G. and Dina, D. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2627–2631.
- [5] Ovchinnikov, Yu.A., Sverdlov, E.D., Tsarev, S.A., Arsenyan, A.G., Rokhlina, T.O., Chizhikov, V.E., Petrov, N.A., Prikhod'ko, G.G., Blinov, V.M., Vasilenko, S.K., Sandakhchiev, L.S., Kusov, Yu.Yu., Grabko, V.I., Fleer, G.P., Balayan, M.S. and Drozdov, S.G. (1985) *Dokl. Akad. Nauk SSSR* 285, 1014–1018 (in Russian).
- [6] Paul, A.V., Tada, H., Von der Helm, K., Wissel, T., Kiehn, R., Wimmer, E. and Deinhardt, F. (1987) *Virus Res.* 8, 153–171.
- [7] Cohen, J.I., Rosenblum, B., Ticehurst, J.R., Daemer, R.J., Feinstone, S.M. and Purcell, R.H. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2497–2501.
- [8] Cohen, J.I., Ticehurst, J.R., Purcell, R.H., Buckler-White, A. and Baroudy, B.M. (1987) *J. Virol.* 61, 50–59.
- [9] Ticehurst, J.R. (1986) *Semin. Liver Dis.* 6, 46–55.
- [10] Lemon, S.M. (1985) *N. Engl. J. Med.* 313, 1059–1067.
- [11] Lemon, S.M., LeDuc, J.W., Binn, L.N., Escajadillo, A. and Ishak, K.D. (1982) *J. Med. Virol.* 10, 25–36.
- [12] Andjaparidze, A.G., Poleschuk, V.F., Zamyatina, N.A., Savinov, A.P., Gavrilovskaya, I.N. and Balayan, M.S. (1985) *Vopr. Virusol.* 4, 468–474 (in Russian).
- [13] Karetnyi, Yu.V., Andjaparidze, A.G., Orlova, T.M. and Balayan, M.S. (1989) *Vopr. Virusol.* 1, 50–53 (in Russian).
- [14] Young, R.A. and Davis, R.W. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1194–1198.
- [15] Sverdlov, E.D., Tsarev, S.A., Markova, S.V., Vasilenko, S.K., Chizhikov, V.E., Petrov, N.A., Kusov, Yu.Yu., Nastashenko, T.A. and Balayan, M.S. (1989) *Life Sci. Adv. Ser. Mol. Genet.*, in press.
- [16] Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499–560.
- [17] Pozdnyakov, V.I. and Pankov, Yu.A. (1981) *Int. J. Peptide Protein Res.* 17, 284–291.
- [18] Bradley, D.W., Schable, C.A., McCaustland, K.A., Cook, E.H., Murphy, B.L., Fields, H.A., Ebert, J.W., Wheeler, C.M. and Maynard, J.E. (1984) *J. Med. Virol.* 14, 373–386.
- [19] Lemon, S.M., Chao, S.-F., Jansen, R.W., Binn, L.N. and LeDuc, J.W. (1987) *J. Virol.* 61, 735–742.