Primary structure of the SSAV tether-RNase H endonuclease (pol) region deleted in SSV

Ruth Brack-Werner, Thomas Werner¹, Christine Leib-Mösch², Rüdiger Hehlmann³ and Volker Erfle¹

GSF-Abteilung für Molekulare Zellpathologie, ¹GSF-Institut für Säugetiergenetik, Ingolstädter Landstraße 1, D-8042 Neuherberg, ²Medizinische Poliklinik der Universität, D-8000 München and ³III Medizinische Klinik Mannheim der Universität Heidelberg, 6800 Mannheim, FRG

Submitted June 5, 1989

EMBL accession no. X15311

The v-sis bearing simian sarcoma virus (SSV) contains a deleted pol gene relating to only the 3' endonuclease section and therefore requires an associated helper virus (SSAV) for replication. Sequence analysis of an SSAV pol region not contained in SSV was carried out with subcloned SSAV pol fragments (map positions 7.8 - 9 and 0 - 0.8 of B11 (1)) derived from the molecular SSAV clone pB11 (provided by E.P. Gelmann). The complete nucleotide sequence given below was assigned to tether, RNase H and part of the endonuclease according to (2). Alignment of the translated tether and RNase H amino acid sequence with the corresponding pol sequences of baboon endogenous virus (BaEV) and murine leukemia virus AKV yielded a slightly higher score for BaEV (identical amino acids underlined in figure). The 5' endonuclease showed only weak similarity (not marked in figure).

The SSAV sequence is related to human endogenous retroviral element S71 as shown by phylogenetic analysis (4). Comparisons of the tether sequences of SSAV, BaEV, AKV and S71 with each other indicates that this region is as well conserved as the RNase H domain. The SSAV sequence between position 1337 and 1431 is identical to the SSV pol sequence between position 2903 and 2997 with a single A/C mismatch at 1348/2914. Together with the SSV endonuclease sequence (3) this overlap enables construction of a 2.2 kb SSAV/SSV sequence encompassing all pol domains downstream of the reverse transcriptase.

Restrictions sites present in the map of (1): PstI (1-6), HindIII (90-104), BamHI (271-276), HpaI (349-353), SalI (1054-1059), EcoRI (1154-1159), HpaI (1177-1182), PstI (1251-1256). Additional restriction site: Sac I (1130-1135). Border: tether/RNase H: 617; RNase H/endonuclease: 1084 (2)

```

CTGACGCTC TGGATCCGCT GGTTCGCTCT CCGGCTGCA CCGTTGACC CCCTAACAA GGAAGCATC CTTTATATC GGACTGAGG ACATCAGG AA GCT TGACC GCATAAAGA 120
CRLWTFGFAS LAAPLYP LTERESI PFIM TERHQK AFDRIKE
AGCGTCTGCT TGAAGCCGCT CTTTGCGCT CCAGAGCTC GGCANACAT CCGTAGATGAG AGCGCTGGG TGCGCCGGGG AGCTCTTACT CAGACTTTAG GACCGTGCG 240
ALLSAPALALPLDL AKFFTL YVDEB RAGV ARGVLTQT LGFWNR
CGCGCCGGTA GCTTATCTAT CGAAGAAGCT GGATCGCTG GCGAGCGGT GCGCAACCTG CCTAAAGCG GTTCAGCAGC TGCGACTCT TCTCAAGGAC GCTGATAT GT TACG TTGGG 360
RFVAYLS KKLDPVAS GMPF TCLKA VAAV ALLLKD ADKLT LG
ACAAATGTTG ACTGTGATG CTTCCATAG CCGTGAAGAC ATCGTGGCGC AGCGCCCGCA CCGGTGGATG ACCAATCCA GGTATGACTA TTACGAGAC CTCTGCTATA ATGAAGGCT 480
QHV TVIA SHS LES IVRQ PPD RWM TNAR MTH YQS LLLN ERV
ATGTTGGCG CCCTCTGGG TCTTGACCT AGCTAACCTA CTACAGTCG AGTCGGAAC TACCCGATG CACAGGTGCT CAGAACTCT CGCGGAAGAA ACTGGAATC GACGAGACT 600
SFA PPAV LNPATL LPEVE SEA TPV HRCSE ILAEET GTTR RDL
GAAGACAG CCAATGCCG GGTCCGACG CTGGTATAG CAGCGTAGCA GTTTCATGCG GGAAGTAAA CGGAGAGAG GGGCCCGCAT CCGTAGATGCG AAGCGAGCG TGTGGCCAG 720
RDQ PLPG VPA WYTDGSS FIA EGK RRAGAAI VDG KRTV HAS
CAGCTTGCCA GAAGTACGT CAGCCAGAA GCGCGACTA GTGCGCTTGA CCGAGCATT AGCGCTGGCG GAAGGAGAG ACATCAACAT CTACACAGAC AGCAGTATG CTTTGGCCAC 840
SLP EGT S AQK AEL VALT QAL RLA EGRD INI YTD SRYA FAT
TGCTCATATT CATGGGCCA TATATAACA GAGGGGGCTG CTCACTCTG CTGGAAGAA CATTAAAC AAAGAGAAA TTTTGCGCT CCGTAGAGCC ATTCACTCT CTAAGCGGT 960
AH I HGA IYKQ RGL ETS A GRD IKNK EEL LALE A IHL P KRV
CGCATATC CACTGCGCG GGCACAGAA GGGAAATGAC CCGTGGCCA CCGGGAACCG GAGGCGCGAC GAGGCTGCAA AACAGCCCG CT GTGCG AGAGTCTGG CAGAACTAC 1080
AII HCPG HQRG ND PVAT GNR RAD EAAK QAA LST RVLA ETT
AAACCTCAA GAGCTATCA ANCCGCTCA GGTTAAGACC AGGCGAGG E AGCT ACCCG TGACCGGGG AAC GAATG TTCAGCGGT ACATC GTTA AG ACACCTTAG GACCAGAGAA 1200
K P O E L I K P A Q V K T R P G E L T P D R G R E F I Q R L H Q L T H L G P E K
GCTTTCGCA CTAGTAACG CGACCGAGCT CCTCATCCCA AACCTCAGT CTGCA TCG CGAAGTCACC AGTCAGTCT AGCGTTGTGC CATGACTAAT GCGGTACAAA CCTACCGAGA 1320
L L Q L V N R T L I P N L Q S A V R E V T S Q C Q A C A M T N A V T T Y R E
GACCGGAAG AGGCACGAG GAGATCGACC CGCGGTGTC TGGAGGTAG ACTTCACAGA AGTGAAGCT GCGCGGTATG GAACACGTA TCTCTGGTA TTCAAGATA C 1431
T G K R Q R G D R P G V Y W E V D F T E V K P G R Y G N R Y L L V F I D

```

REFERENCES:

- (1) Gelmann, E. P., Wong-Staal, F., Kramer, R. A., Gello, R. C. (1981) Proc. Natl. Acad. Sci. USA 78, 3373-3377.
- (2) Johnson, M. S., McClure, M. A., Fang, D.-F., Gray, J., Doolittle, R. F. (1986) Proc. Natl. Acad. Sci. USA 83, 7648-7652
- (3) Devere, S. G., Reddy, E. P., Law, J. D., Robbins, K. C., Aaronson, S. A. (1983) Proc. Natl. Acad. Sci. USA 80, 731-735.
- (4) Werner, T., Brack-Werner, R., Leib-Mösch, C., Backhaus, H., Erfle, V., Hehlmann, R. (submitted)