Molecular Cloning and Complete Nucleotide Sequence of the Attenuated Rabies Virus SAD B19

KARL-KLAUS CONZELMANN,[1] JAMES H. COX, LOTHAR G. SCHNEIDER, AND HEINZ-JÜRGEN THIEL

Federal Research Centre for Virus Diseases of Animals, Paul-Ehrlich-Strasse 28, D 7400 Tübingen, Federal Republic of Germany

Received September 19, 1989; accepted November 22, 1989

Complementary DNA spanning the entire genome of the attenuated rabies virus strain SAD B19 which is used for oral immunization of foxes in Europe was cloned and sequenced. The viral genome comprises 11,928 nucleotides and encodes the five viral proteins N, NS, M, G, and L. Deduced protein sequences are highly similar to those of the pathogenic PV strain, homologies ranging from 90.6% for the M to 98.6% for the L protein. The five cistrons are separated by intergenic regions of 2, 5, 5, and 24 nucleotides, respectively. The G transcription stop/polyadenylation consensus signal in SAD B19 is destroyed by a deletion of three A residues. The strong conservation of both noncoding and coding nucleotide sequences indicates a high selective pressure on the primary structure of rabies virus genomic RNA.

INTRODUCTION

The neuropathogenic rabies virus is propagated in Europe mainly by foxes. Since 1983 a successful field trial has been performed with the SAD B19 rabies virus vaccine strain, which immunizes foxes after oral application. No case of rabies has yet been reported which was caused by this vaccine strain. Safety and efficacy of the SAD B19 live vaccine have led to a drastic decrease of rabies cases and eradication of wildlife rabies, at least in parts of Europe, appears realistic (Schneider and Cox, 1983; Schneider *et al.*, 1988).

The SAD B19 strain originated from a street virus (SAD, Street Alabama Dufferin) isolated from a rabid dog and was attenuated by multiple passages in different cell types. In contrast to the closely related Evelyn–Rokitnicky–Abelseth (ERA) strain, which originated from the same isolate but differs in cell culture history and which has retained a residual pathogenicity, it is nonpathogenic for a wide range of animal species when administered orally or intramuscularly.

Relatively little is known on the molecular biology of rabies virus, especially replication, transcription, and putative recombination. Until recently only nucleotide sequences of the glycoproteins of some laboratory virus strains and short stretches of the 3′-end of the RNA genome were available (Anilionis *et al.*, 1981; Kurilla *et al.*, 1984; Yelverton *et al.*, 1983; Rayssiguier *et al.*, 1986). The first complete nucleotide sequence of a rabies virus genome, the highly pathogenic strain PV, was published recently (Tordo *et al.*, 1988).

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. M31046.

[1] To whom requests for reprints should be addressed.

Rabies virus is a member of the family Rhabdoviridae, genus *Lyssavirus,* which possess a negative-stranded RNA genome of about 12 kb. The genome is organized similar to that of the better characterized vesicular stomatitis virus, genus *Vesiculovirus,* and encodes five major proteins (Coslett *et al.*, 1980; Holloway and Obijeski, 1980). As in the vesiculoviruses the N, NS, and L proteins together with the genomic RNA form a nucleocapsid which is enveloped by a membrane containing the transmembrane glycoprotein G, whose spikes are mainly responsible for the antigenicity of the virus. The M protein is located at the inner side of the membrane, providing a connection between the nucleocapsid and the cytoplasmic domain of the glycoprotein (Cox *et al.*, 1981). The amino acid sequences of the isofunctional VSV and rabies virus proteins, however, are highly dissimilar and only the nucleoproteins N and the polymerases L display low, segmented homology. The existence of a pseudogene region (Ψ) between the G and L cistrons and the heterogeneity of the intergenic regions in rabies virus has further distinguished rabies virus and VSV (Tordo *et al.*, 1986b, 1988).

The knowledge of nucleotide and protein sequences of rabies viruses may lead to an understanding of the biological differences between *Lyssa-* and *Vesiculoviruses.* In addition, comparison of highly pathogenic rabies virus strains like PV with attenuated strains can provide insight into pathogenicity mechanisms of this important disease. In order to characterize the rabies virus vaccine strain SAD B19 on a molecular basis we undertook molecular cloning and sequencing of its RNA genome and compared it to that of the pathogenic rabies virus strain PV.

MATERIALS AND METHODS

Cells and virus

BHK-21 cells (clone BSR) were infected at an m.o.i. of 0.1. After 30 min adsorption at room temperature, cells were washed twice and incubated at 37° in Eagle's medium supplemented with 0.3% BSA in a 5% CO_2 atmosphere. Virus was purified as described (Wiktor *et al.*, 1977).

Oligonucleotides

Deoxyoligonucleotides were synthesized on a Biosearch 8700 DNA Synthesizer and purified by denaturing polyacrylamide electrophoresis. The oligonucleotides used in this study are ol 18 (5′-pdTTTAGCGACCGTTCGATC-3′); ol 17 (Fig. 2, position 149; 5′-pdGAGTACAAGTACCCTGCC-3′); ol 19 (Fig. 2, position 11723; 5′-dCATAGGTACAACAGGTG-3′); and ol 40 (5′-pdGATCGATCGAATTCATATGCGCGATCGAACGGTCGCTAAA-3′).

5′ Labeling of oligonucleotides

Labeling was carried out with 50 ng of oligonucleotides in 30 μl of 70 m*M* Tris, pH 7.6, 10 m*M* $MgCl_2$, 5 m*M* DTT, 4 U T4 PNK, and 50 μCi [γ-^{32}P]ATP (3000 Ci/mmol; Amersham). Hybridization to Northern filters was performed at 45° for 14 hr in 1 *M* NaCl, 0.1 *M* Tris, 20 m*M* sodium phosphate, 1 μ*M* EDTA, 0.5% SDS, 10× Denhardt's solution, and 50 μg of tRNA per milliliter. The filters were washed three times at 45° in 5× SSC, 0.1% SDS for 20 min each, and exposed to Kodak X-Omat AR films at −70° with an intensifying screen (Agfa Curix MR 800).

Oligonucleotide–RNA ligation

Forty nanograms of the synthetic oligonucleotide ol 40 were ligated to 4 μg of genomic RNA with 150 ng of T4-RNA ligase (Pharmacia) in 30% PEG, 50 m*M* Tris–HCl, pH 7.8, 5 m*M* $MgCl_2$, 0.5 m*M* ATP, 10 m*M* β-mercaptoethanol for 2 hr at 20° in a volume of 10 μl.

Molecular cloning

Genomic RNA was isolated according to Chirgwin *et al.* (1979) and checked for integrity by denaturing agarose gel electrophoresis (McMaster and Carmichael, 1977). cDNA synthesis was performed as described by Gubler and Hofmann (1983) using the synthetic primer ol 17. cDNA molecules larger than 3 kb were isolated by preparative agarose gel electrophoresis, ligated with synthetic *Eco*RI adaptors (Pharmacia) and cloned in λgt10 phages (Promega). SAD B19-specific clones were identified by Northern hybridization with total RNA of rabies-infected cells.

cDNA containing the 3′ end of the genomic RNA was made after ligation of ol 40 to the 3′ end of genomic RNA by priming with ol 18 and cloned in λZAP II phages (Stratagene).

Sequence determination and analysis

Subclones of positive λgt10 inserts in pEMBL18 and -19 phagemids (Dente *et al.*, 1983) were subjected to unidirectional deletion using exonuclease III and S1-nuclease according to Henikoff (1984). Both strands of deletion clones were sequenced after isolation of single-stranded DNA by the chain termination method (Sanger *et al.*, 1977) using T7-DNA polymerase. pBluescript plasmids were excised from λZAP II phages *in vivo* according to the supplier's instructions and were sequenced by double strand plasmid sequencing according to Zhang *et al.* (1988). The last 30 nucleotides at the 5′ end of the genome were determined by sequencing purified genomic RNA using the synthetic primer ol 19 which was deduced from the sequence obtained from the deletion clones (Zimmern and Kaesberg, 1978).

Computer analysis of the nucleotide and peptide sequences was performed using the UWGCG software (Devereux *et al.*, 1984) on a Micro VAX II (Digital).

RESULTS AND DISCUSSION

Molecular cloning and sequence determination of the SAD B19 genome

For the preparation of a cDNA primer a sequence stretch of the PV N gene was chosen which is identical to the sequence of challenge virus strain (CVS) and shows a high homology to the corresponding region of the VSV N gene (Tordo *et al.*, 1986a; Kurilla *et al.*, 1984; see also Fig. 2, position 149 to 166). The ability of the oligonucleotide ol 17 to prime cDNA synthesis of the SAD B19 genome was confirmed by 5′ labeling and Northern hybridization to SAD B19 genomic RNA. About 80% of the obtained λgt10 cDNA clones proved to be specific for SAD B19 by hybridization to total RNA of infected cells. The other clones are likely to result from residual cellular RNA present in the RNA preparations, mainly ribosomal RNAs. Besides the correctly primed clones we found a considerable amount of SAD B19-specific clones that obviously resulted from either random priming of oligonucleotides or selfpriming of the genomic RNA. The location of positive clones used for further analyses is shown in Fig. 1.

In order to clone the 3′ extremity of the genome the synthetic oligonucleotide ol 40 was ligated to the 3′ end of the genomic RNA. cDNA synthesis was primed by oligonucleotide ol 18 which is complementary to the extreme 3′ end of ol 40. The resulting cDNA was cut with *Eco*RI (the first site is provided by the sequence of

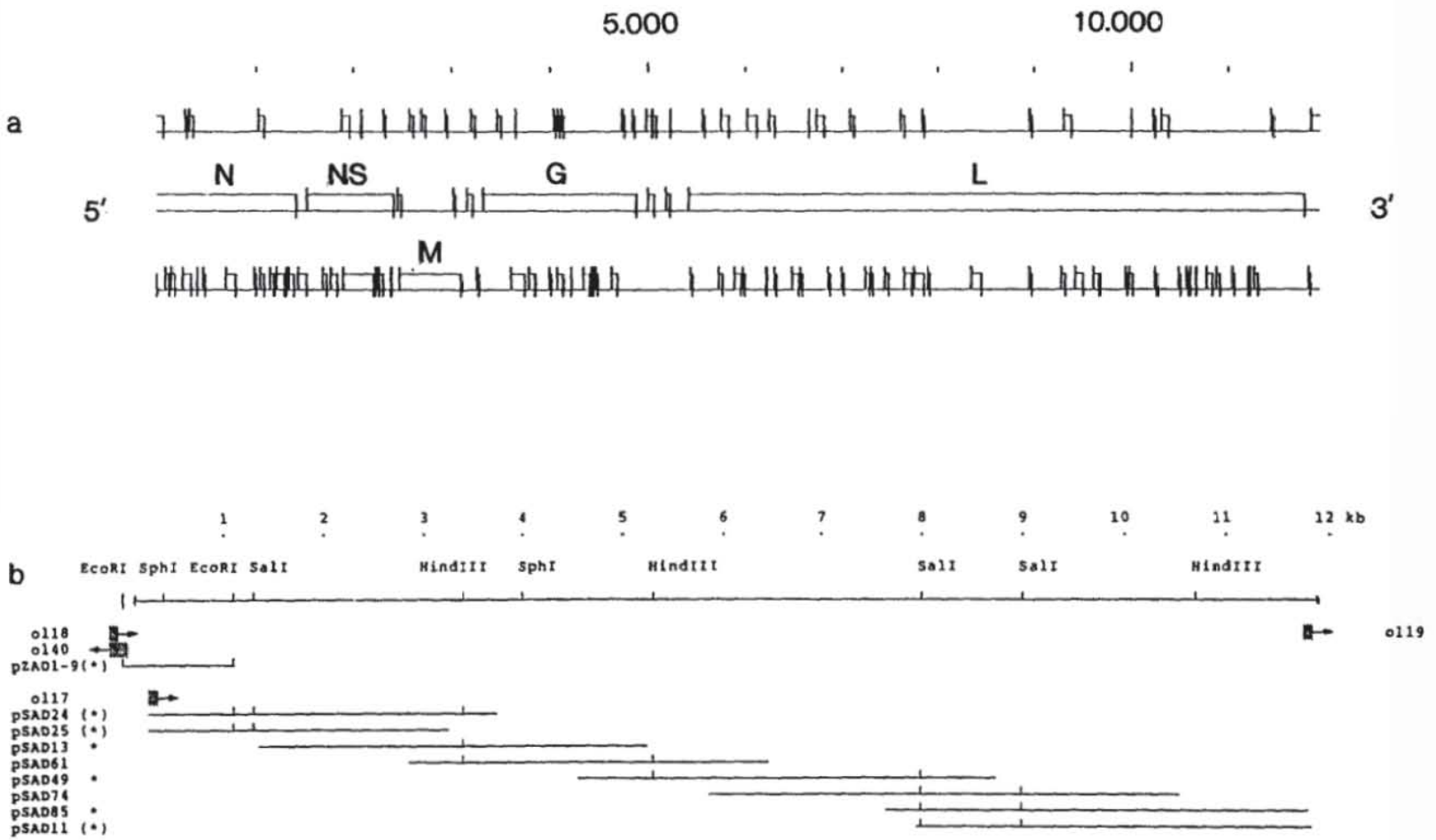


FIG. 1. (a) Open reading frames of the SAD B19 positive strand RNA. (b) Localization of oligonucleotide primers used for cDNA synthesis or genomic RNA sequencing and of cDNA clones which were sequenced entirely * or partially (*). ol 17, cDNA primer; ol 19, RNA sequencing primer; ol 40, oligonucleotide which was ligated to the 3′ end of genomic RNA; ol 18, cDNA primer complementary to ol 40; pSAD, λgt10/pEMBL clones; pZAD, λZAP II/pBluescript clones.

the 40-mer-nucleotide, the second is the only *Eco*RI site of the SAD B19 genomic sequence, located 1.1 kb from the 3′ end) and cloned in λZAP II phages. Clones with inserts representing the 3′ end of the SAD B19 genome were identified by hybridization with a 200-base *Eco*RI/*Sph*I fragment of clone pSAD 24. In Fig. 1 the cloning and sequencing strategies are summarized and the location of the oligonucleotide primers used for priming cDNA synthesis and genomic RNA sequencing are shown.

Nucleotide and deduced amino acid sequence analysis

We have determined the nucleotide sequence of the entire SAD B19 genome (Fig. 2) except for the very last nucleotide at the 5′ end, which cannot be analyzed by the enzymatic sequencing protocol. The SAD B19 genome comprises 11,928 nucleotides, and is thus 4 nucleotides shorter than the genome of the PV strain (Tordo *et al.*, 1986a,b, 1988). With respect to the PV genome one A residue and a block of three A residues are deleted in the 3′ noncoding regions of the M and G genes, respectively (Fig. 2).

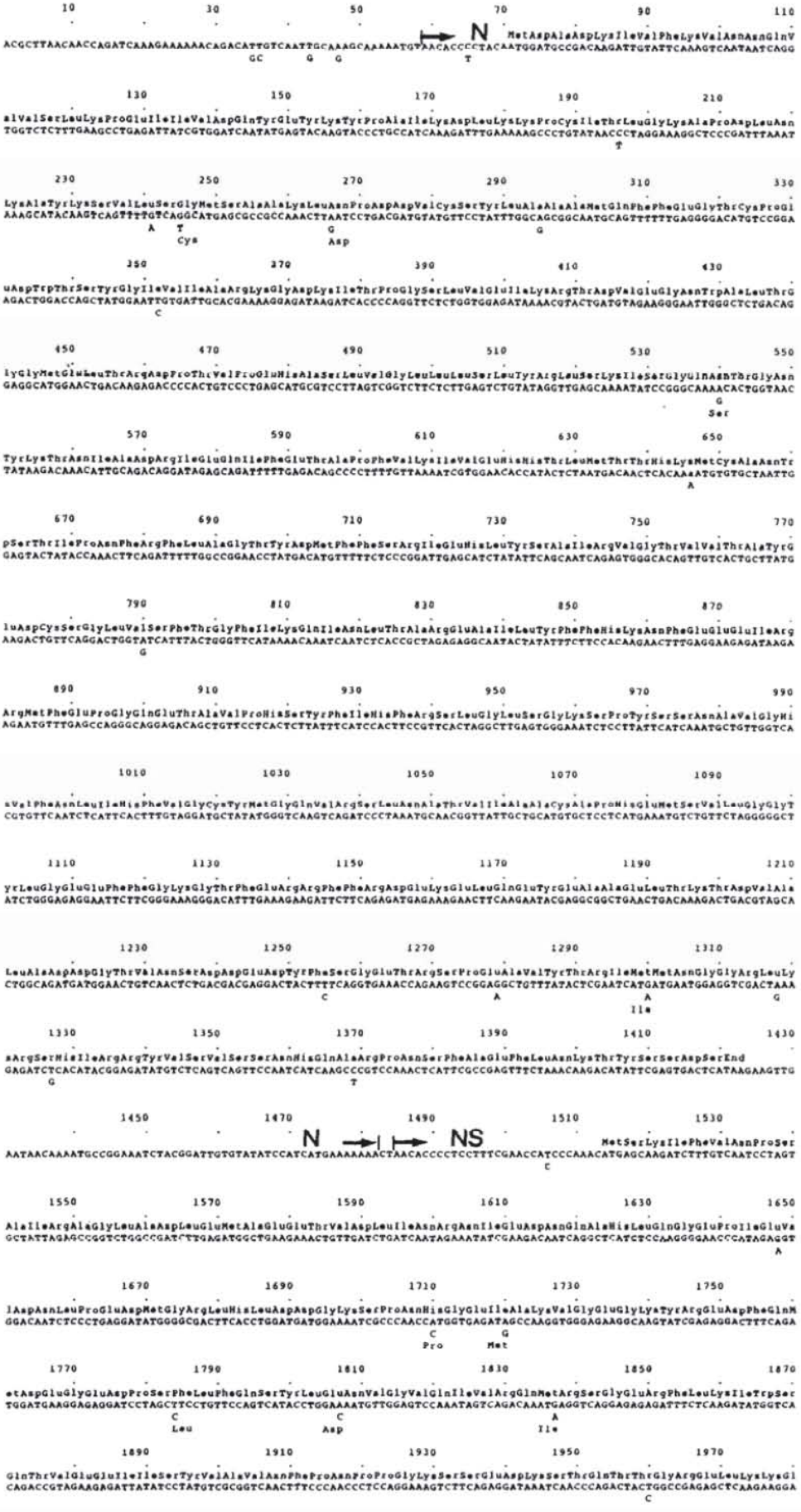
1. Open reading frames. The genome organization of SAD B19 is identical to that of the PV strain. Five major open reading frames (ORFs) encode the viral N, NS, M, G, and L proteins. As in the PV genome an additional ORF within the coding region of the NS gene is found both in SAD B19 and ERA (unpublished results) which could encode a protein of 102 amino acids length (Fig. 1). The NS gene sequences of all vesiculoviruses also possess a second open reading frame which could encode proteins of 65 or 67 amino acids (Lazzarini, cited in Banerjee, 1987). Actually the phosphoprotein genes of VSV and different paramyxoviruses, which are more distantly related to rhabdoviruses, encode additional polypeptides (Giorgi *et al.*, 1983; Bellini *et al.*, 1985; Barrett *et al.*, 1985; Luk *et al.*, 1986; Herman, 1987). No evidence for expression of the second NS ORF by SAD B19 or ERA was found up to now. In the pseudogene region, supposed to represent a remnant gene (Tordo *et al.*, 1986b), the SAD B19 possesses similar to PV only short ORFs, capable of encoding 17 and 18 amino acids, respectively. The negative strand RNA of SAD B19 shows five ORFs with a coding capacity of more than 100 amino acids (not shown). No evidence for the expression of these ORFs has been found.

2. Coding regions and proteins: (a) N gene. The nucleoprotein gene of SAD B19 encodes 450 amino acids. The deduced amino acid sequence differs in four residues from that of the PV strain (99.1% homology). Three of the substitutions are located in the amino terminal moiety of the protein and one in the carboxy terminal region. The recently corrected version of the ERA N Protein (Ertl *et al.*, 1989) contains two different amino acids compared to the SAD B19 sequence (99.6% homology). None of the PV or ERA substitutions is located within the four particularly conserved regions found when PV and VSV Indiana N protein sequences are compared (Tordo *et al.*, 1986a). One of these regions (aa 225 to 247) overlaps with a stretch of 34 aa which is conserved among vesiculoviral N and paramyxoviral NP proteins. Another region which is highly conserved in vesiculoviruses shows 16 identical amino acids out of 31 (Masters and Banerjee, 1987). The corresponding region too is identical in the PV and SAD B19 proteins. The putative casein-type phosphorylation site (Ser, 389) (Dietzschold *et al.*, 1987) is also not affected. As in other rhabdoviral N proteins the carboxy terminus contains a large number of highly charged basic amino acids and may represent the RNA binding site of the protein (Gilmore and Leong, 1988). One or both of the above-described conserved sequence stretches might also be involved in RNA binding.

(b) NS gene. The amino acid sequences of the SAD B19 and PV phosphoproteins are well conserved and show 97.3% homology. Considering only nonconservative amino acid replacements the homology increases to 99.3%. Especially the highly hydrophilic regions of the protein are conserved. There is no amino acid replacement in the large central hydrophilic region of the protein (aa 139 to 170) or the shorter hydrophilic stretches in the amino terminal third of the protein. The amino acid replacement at position 177 substitutes a potential phosphate acceptor Thr. In contrast to the vesiculoviral NS proteins where the highest degree of hydrophilicity and most of the phosphate residues are found in the amino terminal part of the protein (Masters and Banerjee, 1987; Gill and Banerjee, 1985; Bell and Prevec, 1985), the most hydrophilic region in rabies virus NS is in the center of the protein.

When the five viral proteins of VSV serotypes (New Jersey, Indiana) and Chandipura virus are compared, the NS protein shows the lowest degree of homology (minimum 21%; Masters and Banerjee, 1987),

FIG. 2. Total genomic sequence (line 2) and translation products (line 1) of SAD B19 compared to those of the PV strain (lines 3 and 4, respectively). In the PV sequences only deviations from SAD B19 are shown. Arrows indicate transcription start and stop signals. Numbers above the sequences correspond to SAD B19 nucleotides, numbers at the right side to amino acids. The PV genome possesses 4 additional nucleotides as shown in the PV sequence.


N
MetAspAlaAspLysIleValPheLysValAsnAsnGlnV
N
NS
MetSerLysIlePheValAsnProSer

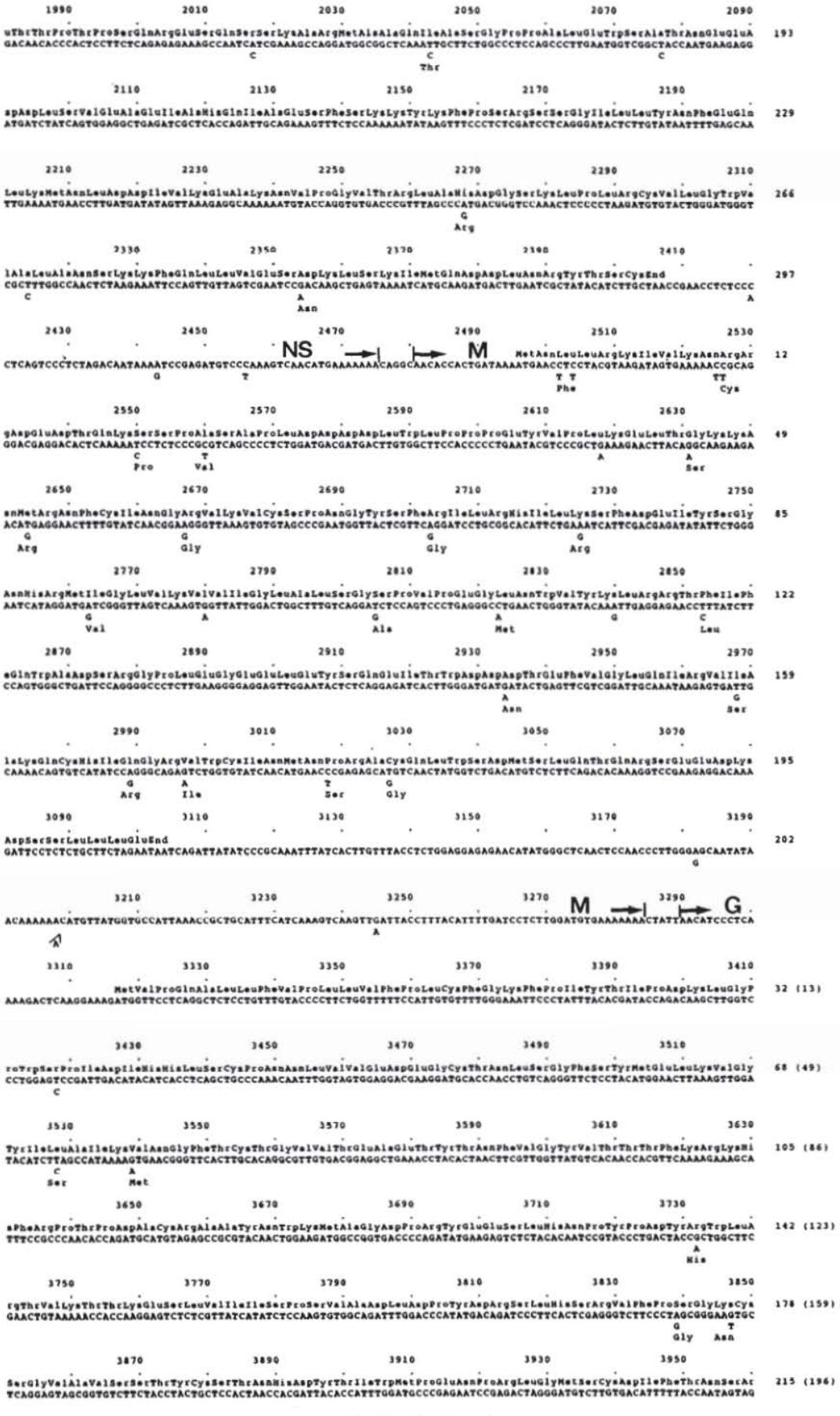


FIG. 2—*Continued*

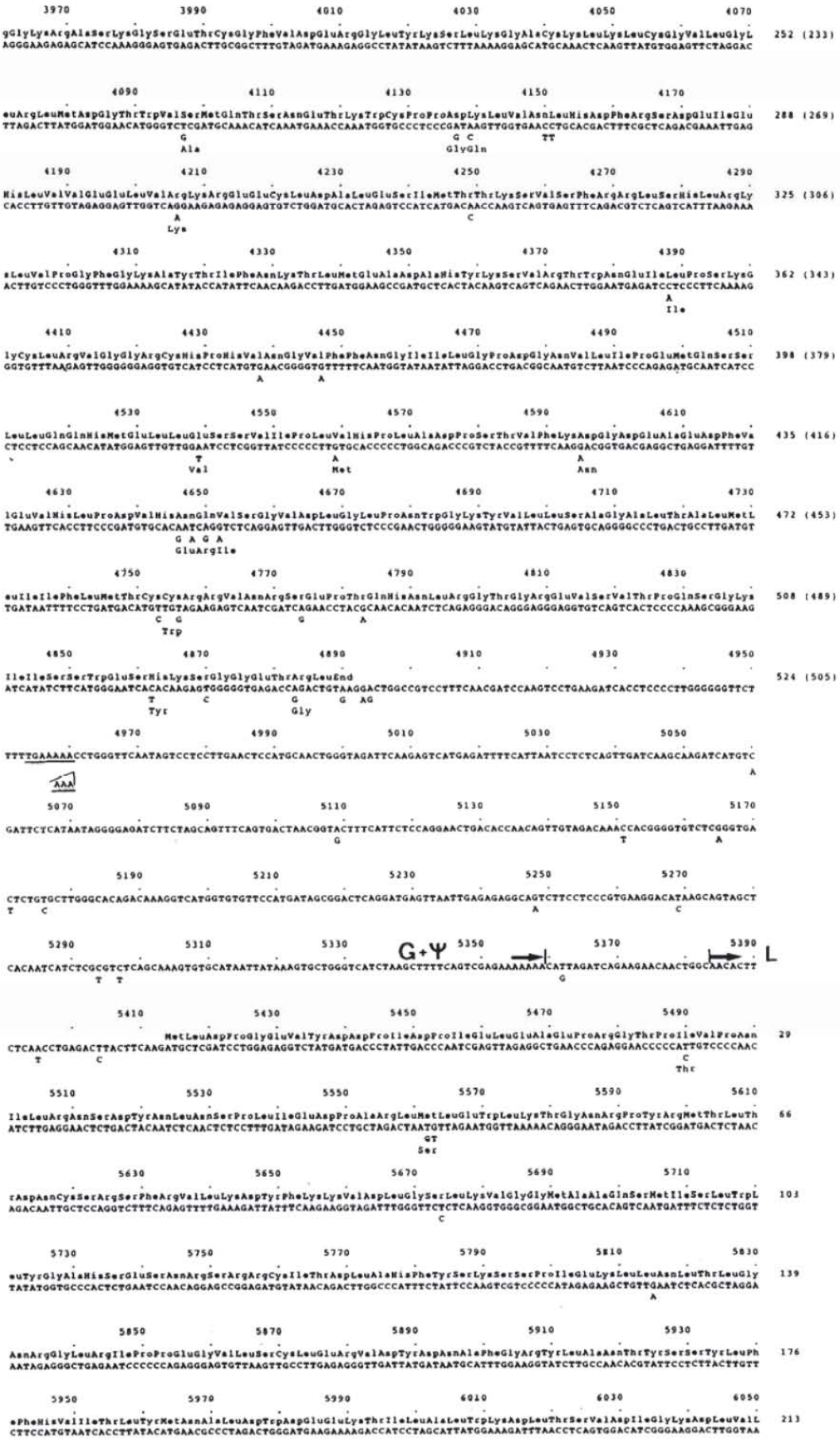


FIG. 2—*Continued*

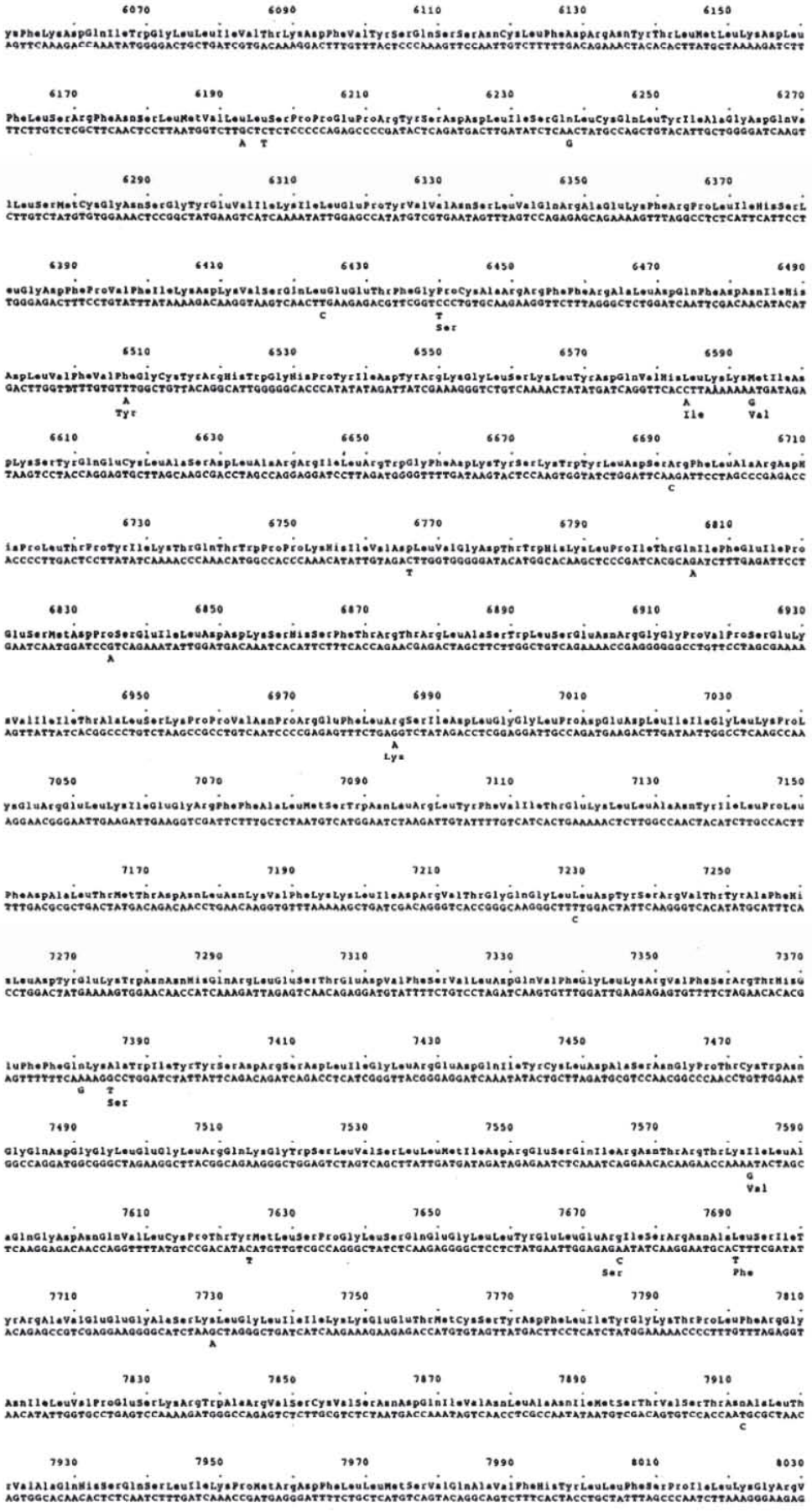


FIG. 2—*Continued*

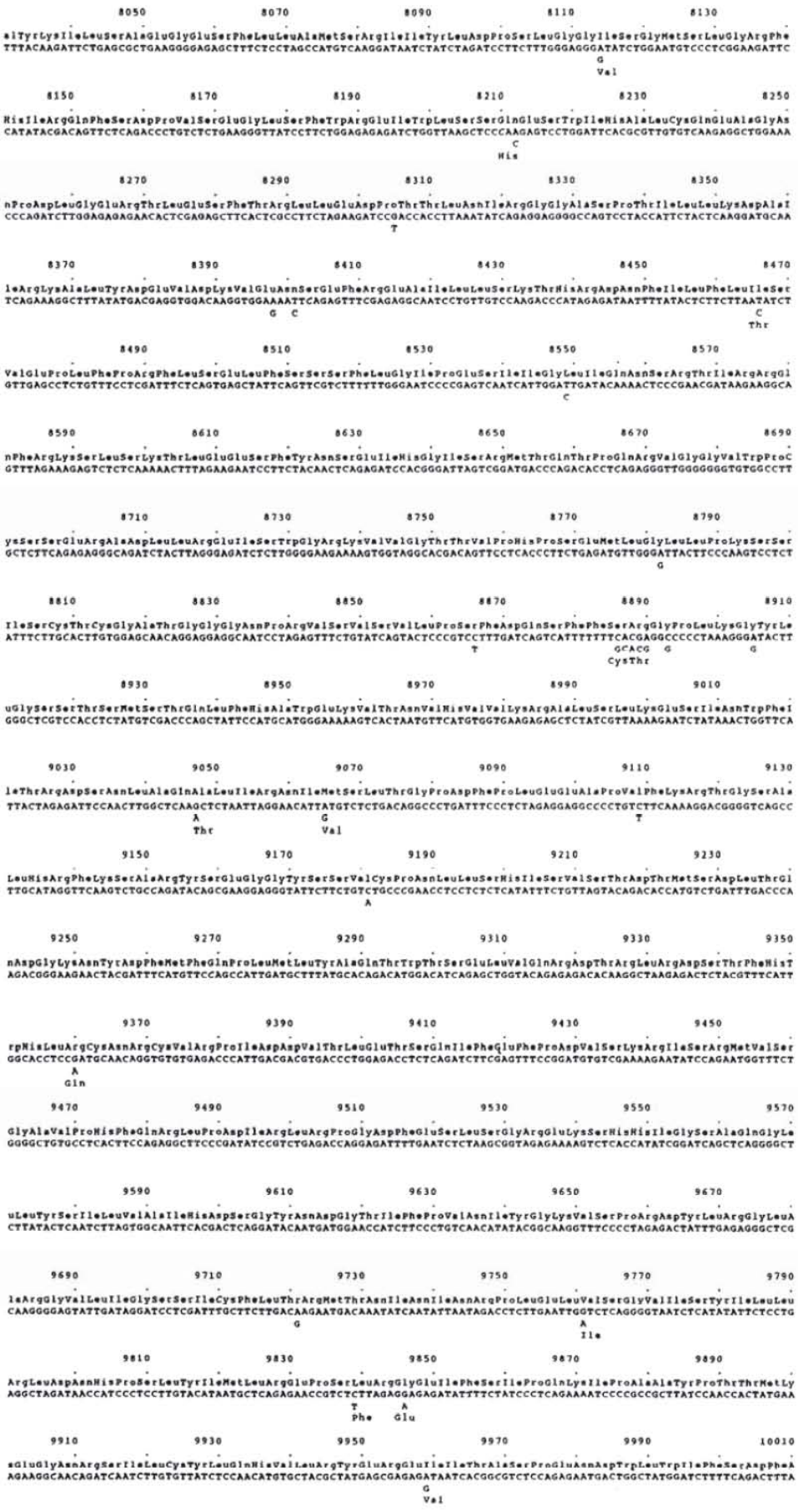


FIG. 2—*Continued*

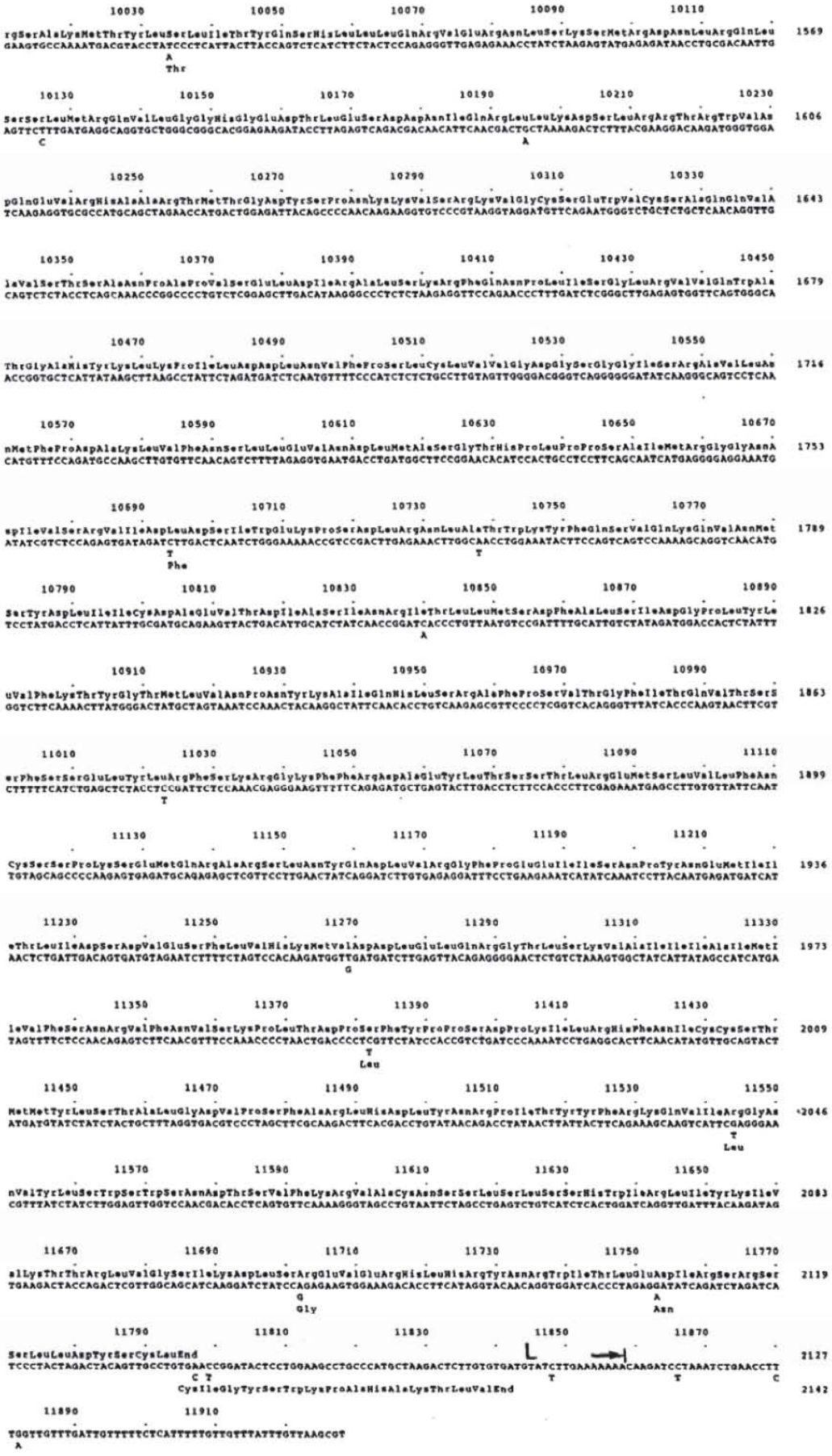


FIG. 2—*Continued*

whereas the rabies NS protein is much more conserved than the M protein (see below). Similar to vesiculoviruses, paramyxoviruses exhibit also a larger evolutionary divergence between the phosphoproteins than between the matrix proteins (Bellini *et al.*, 1986). It remains to be determined whether the results of the comparison of NS and M proteins from rabies virus strains SAD B19 and PV apply to the whole *Lyssavirus* genus.

(c) M gene. The M gene shows 19 amino acid replacements out of 202 compared to the PV sequence (90.6% homology) and is thus the most variable rabies virus-encoded protein. The basic features found in the PV M protein, i.e., hydrophilic regions at the amino and carboxy terminal regions and a hydrophobic putative membrane bound region in the center of the protein, are conserved in SAD B19.

A relatively low homology of 96% (eight replacements) is also found between the SAD B19 and the closely related ERA strain M proteins (Rayssiguier *et al.*, 1986) and compared to the other known ERA proteins the M also displays the highest variability.

It has been shown for temperature-sensitive mutants of a paramyxovirus, Newcastle disease virus, that mutations in the M gene have a drastic influence on incorporation of the surface glycoprotein F, which led to a decreased infectivity (Peeples and Bratt, 1984). With regard to attenuation of rabies virus changes in the matrix protein may play an important role. It is conceivable that the conformation of the envelope glycoprotein is modified by its interaction with an altered matrix protein. Mutations in the M gene also contribute to the establishment of measles virus persistence, resulting in human subacute sclerosing panencephalitis (Cattaneo *et al.*, 1986, 1988). It will be important to investigate the functions of rabies virus M protein, especially its interactions with both the glycoprotein and the nucleocapsid.

(d) G gene. The transmembrane glycoprotein G which represents the major surface antigen of the virus, organized in peplomers on the membrane surface, is encoded by an ORF of 1575 nucleotides (524 amino acids, including the signal sequence) like all glycoproteins of rabies virus strains analyzed so far (Tordo *et al.*, 1986b; Anilionis *et al.*, 1981; Yelverton *et al.*, 1983; Prehaud *et al.*, 1988). The overall amino acid homologies of these glycoproteins range from 88.2% (SAD B19/CVS) to 99.4% (SAD B19/ERA). The closely related ERA and SAD B19 proteins show 3 amino acid replacements. The PV glycoprotein differs from the ERA and SAD B19 proteins by 16 and 19 amino acid exchanges (97.0 and 96.4% homology, respectively).

The mature glycoprotein of rabies virus and also VSV consists of an ectodomain or "antigenic domain" present on the surface of the virion, a transmembrane domain, and a cytoplasmic domain located within the viral envelope (Fig. 3). The deduced signal sequences (aa −19 to −1) of the SAD B19, ERA, and PV glycoproteins are identical and show four amino acid replacements compared to CVS. Interestingly, the transmembrane area (aa 440 to 461) and the cytoplasmic region (aa 462 to 505) are well conserved in the SAD B19, ERA, and PV strains with one and two replacements, respectively, whereas the CVS sequence of these domains diverges extremely from that of SAD B19 or PV (63.6% homology with SAD B19 and ERA and 62.1% with PV).

The ectodomain (aa 1–439) of the SAD B19 protein shows 16 differences to that of the PV protein. The most divergent region of the ectodomain lies close to the transmembrane segment with six exchanges located between aa 388 and 428.

Three of the four potential *N*-glycosylation sites of the PV ectodomain are conserved in the SAD B19 and also in the ERA strain (37, 247, 319). The glycosylation site at residue 158 of the PV protein may be specific for this strain, since both SAD B19 and ERA lack this potential glycosylation site.

The deduced SAD B19 and ERA glycoproteins differ only in residues 56 (Val/Met), 242 (Ser/Ala), and 256 (Lys/Gln), respectively. The biological differences of these strains, especially the enhanced pathogenicity of the ERA strain, might indicate a biological significance of one (or more) of these three amino acid replacements in the glycoprotein ectodomain, which is the major site for virus/cell interaction. While the Val at position 56 is shared by SAD B19 and CVS the residues 242 (Ser) and 256 (Lys) are specific for SAD B19 and are conserved in the other strains (Ala and Gln, respectively). These two replacements are located close to a putative *N*-glycosylation site (247) which is likely to be glycosylated in the ERA strain (Wunner *et al.*, 1985). The exchanges in the SAD B19 glycoprotein might influence the local conformation of the protein and thereby modify the glycosylation of this site. This proposition is supported by the finding that in a CVS-11 mutant with a single amino acid replacement located 6 residues upstream of a glycosylation site only completely glycosylated proteins (i.e., proteins with two glycosylated sites, GI) are found, while in the parent strain two glycoprotein forms are present (GII with only one site, GI with two sites glycosylated; Wunner *et al.*, 1985). Comparative experiments on the glycosylation of the SAD B19 and ERA proteins are necessary to verify this hypothesis.

The exchange of a single amino acid in the antigenic site III (aa 330 to 357) of fixed strains of rabies virus glycoprotein may have significant effects on its biological behavior. By the replacement of Arg (333) with Ile, Glu, or Gln the interaction of the virus with specific neuronal cells is impaired and a significant loss of pathogenicity was observed (Dietzschold *et al.*, 1983; Seif *et*

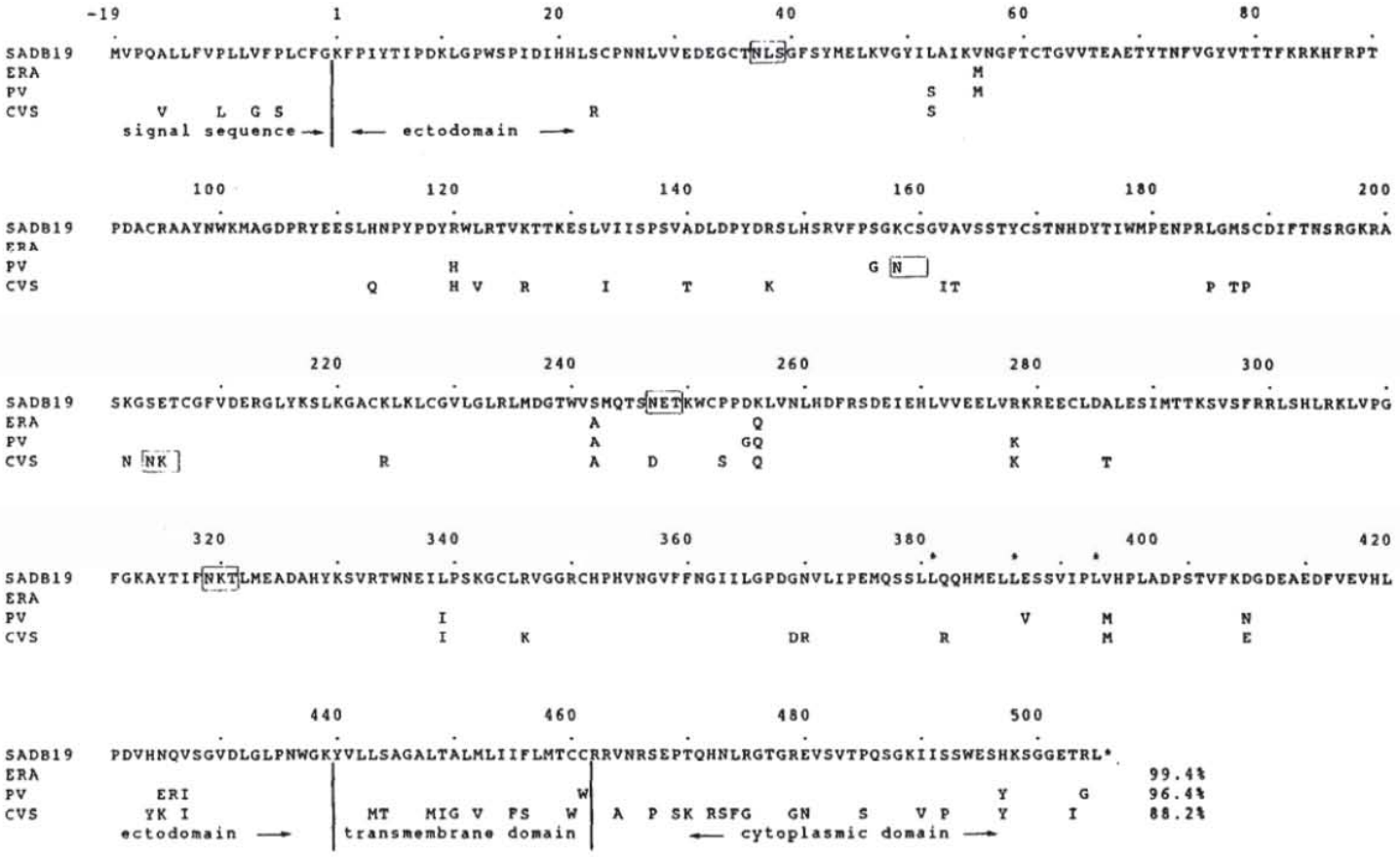


FIG. 3. Comparison of glycoprotein sequences of different rabies virus strains. For the ERA (Anilionis *et al.*, 1981), PV (Tordo *et al.*, 1986b), and CVS strain (Yelverton *et al.*, 1983) only differences to the SAD B19 strain are shown. Homologies to the SAD B19 protein are indicated at the end of the sequences. Potential *N*-glycosylation sites are boxed and the leucine zipper-like motif is indicated by asterisks.

al., 1985; Kucera *et al.*, 1985). The SAD B19 and the ERA strain have identical antigenic sites III. Compared to CVS and PV, however, there is one amino acid replacement at position 339 (Leu in ERA and SAD B19 and Ile in CVS and PV). Both PV and CVS are highly pathogenic strains. Whether the conservative replacement of Ile to Leu at this distinguished site may have a significant influence on the pathogenicity degree of rabies virus cannot be determined at the moment.

As shown by revertants, single amino acid substitutions in regions other than known antigenic sites can influence the biological behavior of rabies virus. Replacement of Leu (132) in the CVS strain by Phe results in temperature sensitivity and avirulence (Prehaud *et al.*, 1989). Whether the differences in the pathogenicity of ERA and SAD B19 which possess an identical antigenic site III and an Arg at position 333 are caused by one or more of such amino acid replacements remains to be investigated.

Sequences resembling a leucine zipper motif which is found in Fos and Jun transforming proteins (Landschulz *et al.*, 1988) and also in the fusion proteins (F) of paramyxoviruses (Buckland and Wild, 1989) are found in the glycoproteins of rabies virus. In the F proteins of paramyxoviruses these motifs are situated close to the transmembrane areas (4 to 11 aa upstream) and are assumed to mediate oligomerization of those proteins (Buckland and Wild, 1989). In the SAD B19 glycoprotein sequence a stretch of three heptadic leucine repeats is found 45 residues upstream of the transmembrane area. This region shows besides the conserved Leu residues a stretch of six amino acids which are identical in the measles virus F protein leucine zipper (Fig. 3). In the glycoprotein of VSV Indiana (Rose and Gallione, 1981) also a leucine zipper-like motif containing three heptadic Leu residues is found about 10 amino acids upstream of the transmembrane region. The location of these motifs in all cases next to the transmembrane area and the lack of heptadic repeats of more than two Leu in other parts of the proteins might suggest a functional significance in peplomer formation.

(e) L gene. The coding region of the L gene of SAD B19 comprises 6348 bases and thus predicts a protein of 2127 amino acids. Surprisingly the deduced PV L protein (Tordo *et al.*, 1988) is 15 amino acids longer than that of the SAD B19. Instead of the stop codon TGA at position 11,795 of the SAD B19 genome the

PV RNA encodes Cys by the codon TGC. In order to exclude the possibility of a cloning artifact in clone pSAD85 two additional cDNA clones spanning the corresponding region were sequenced and the TGA was verified. In addition we sequenced genomic RNA of the SAD B19 and also the ERA strain using the primer ol 19. Both in SAD B19 and also in ERA the existence of the stop codon at position 11,795 was confirmed. The resulting 3′ nontranslated sequence (trailer) of the L mRNA both in ERA and SAD B19 (100% homology) is 61 bases long (including the polyadenylation signal) in contrast to the extremely short trailer of the PV L mRNA which consists of only 16 nucleotides.

The amino acid homology of the SAD B19 and PV L proteins is 98.6% (29 replacements in 2127 residues, not considered the predicted 15 additional amino acids of the PV protein). The L protein of rhabdoviruses is a multifunctional enzyme which is responsible for at least polymerization, capping, and polyadenylation of viral RNAs (Banerjee, 1987). The multiple, most probably independent functional domains of the protein have not been characterized yet. Sequence stretches which show homologies to the VSV and paramyxoviral L proteins and which are assumed to represent conserved catalytic sites have been identified mainly in the central parts of the molecule (Tordo *et al.*, 1988). It is probable, however, that a loss or an addition of 15 amino acids also at the carboxy terminus of this highly conserved protein may have biological effects. Nucleotide and protein sequence information from additional strains and comparative experiments will be needed for the determination of a rabies consensus L carboxy end and its function.

3. Noncoding sequences: (a) RNA termini. The sequence of the genomic 3′ end, which is transcribed to the nontranslated positive 55- to 58-base-long leader RNA in rabies virus (Kurilla *et al.*, 1984), is identical up to position 34 to that of PV. The first 11 nucleotides of the 3′ end of the genome are precisely complementary to the 5′ terminal nucleotides as it was found in PV. The 5′ ends of the two rabies virus genomes are highly conserved, the last 46 nucleotides are also identical; between these and the putative polyadenylation site of the L gene (see above) only 3 nucleotide exchanges are found. The highly conserved structure of the terminal nucleotides demonstrate their significance in the putative function of providing the recognition and initiation site for the viral polymerase complex.

(b) Consensus sequences and intergenic regions. The five protein coding sequences are flanked by conserved sequence motifs, supposed to represent transcriptional start and stop/polyadenylation signals (Tordo *et al.*, 1986b, 1988). Figure 4 shows a comparison of these sequences with those of the PV strain. A striking difference is found in the stop/polyadenylation signal of the glycoprotein mRNA. Compared to PV the SAD B19 genome shows a deletion of three A residues. The resulting motif $TG(A)_5$, however, is also found in coding regions of the N, M, and L genes (positions 177, 2518, and 7117, respectively) and thus we hypothesized that this motif may not function as a transcriptional stop and polyadenylation signal. This theory was confirmed by experiments which showed that transcription of the G mRNA is terminated at the pseudogene stop signal $AG(A)_7$ (position 5351) and not at the modified G "stop/polyadenylation signal" (unpublished results). Thus the pseudogene of SAD B19 does not represent a large intergenic region but a transcribed part of the elongated G cistron. The resulting intergenic region between the G and L cistrons is 24 nucleotides long and starts with a C residue, as do all intergenic regions of rhabdoviruses.

In the PV strain and also in the ERA strain the pseudogene is transcribed, too. (Tordo, personal communication; unpublished results). The stop signals of the PV and ERA G genes possess 8 A residues in contrast to the consensus sequence $TG(A)_7$. This difference probably causes a partial transcription termination resulting in two polyadenylated G mRNAs of 1.7 and 2.3 kb length, the shorter containing only the G-, the other containing the G- and the pseudogene (Tordo, personal communication; unpublished results). An addition of A residues apparently affects the efficiency of the stop signal and for perfect signal function a stretch of exactly 7 A residues may be required.

A rabies virus strain transcribing only the small G mRNA has not been found so far. This finding indicates a selective pressure to delete the stop signal downstream the G coding region and to transcribe the pseudogene to form a fused G/pseudogene mRNA in rabies virus. This might be a step forward to reduce intergene sequences to an essential minimum similar to VSV and Sendai virus where each intergenic region consists of 2 or 1 nucleotide respectively (Rose, 1980; Gupta and Kingsbury, 1984). A close distance from stop to start signal is apparently more favorable for transcription initiation by the viral RNA-dependent RNA polymerase. It is not clear yet whether the enzyme scans the nontranscribed large G/L intergenic region in rabies viruses with an intact G stop signal to the L start signal or whether reinitiation occurs at the L start site.

Both in PV and SAD B19 the functional stop/polyadenylation signal of the pseudogene is different by one nucleotide from the consensus signal $TG(A)_7$ and reads $AG(A)_7$. Thus a less stringent consensus sequence $^{T}/_{A}$-$G(A)_7$ followed by an intergenic sequence beginning with C (positive strand) can account for termination of transcription and polyadenylation.

The intergenic regions, transcriptional stop and start

FIG. 4. Comparison of the SAD B19 and PV transcriptional start and stop/polyadenylation signals and intergenic regions. Differences in the nucleotide sequences are boxed. Ψ = pseudogene.

signals of the N, NS, M, and the stop signal of the L gene are identical in SAD B19 and PV and the nontranslated sequences of the genes show homologies in the range of 91.7 to 100%. The pseudogene sequences are also well conserved (97.4%). This is consistent with the finding of an extremely high nucleotide homology in the protein coding regions of the genome by the usage of identical codons and a rare use of wobble codons to specify an amino acid. Actually the nucleotide homology mostly exceeds the amino acid homology (NS 98.6/97.3, M 96.1/90.6, G 98.2/96.4, and L 99.0/98.6: % nucleotide/amino acid homology, respectively). The only exception is the N gene (99.0% nucleotide and 99.1% amino acid homology). This high pressure on the primary structure of the complete viral RNA suggests its functional role, for example, to provide a favorable template for binding the viral nucleocapsid proteins and/or for optimal transcription and replication.

ACKNOWLEDGMENTS

We are grateful to V. Schlatt for perfect technical assistance and S. Meyer and M. Schwarz for purification of virions.

REFERENCES

ANILIONIS, A., WUNNER, W. H., and CURTIS, P. J. (1981). Structure of the glycoprotein gene in rabies virus. *Nature (London)* **294,** 275–278.

BANERJEE, A. K. (1987). Transcription and replication of rhabdoviruses. *Microbiol. Rev.* **51,** 66–87.

BARRETT, T., SHRIMPTON, S. B., and RUSSELL, S. E. H. (1985). Nucleotide sequence of the entire protein coding region of canine distemper virus polymerase associated (P) protein mRNA. *Virus Res.* **3,** 367–372.

BELL, J. C., and PREVEC, L. (1985). Phosphorylation sites on phosphoprotein NS of vesicular stomatitis virus. *J. Virol.* **54,** 697–702.

BELLINI, W. J., ENGLUND, G., RICHARDSON, C. D., ROZENBLATT, S., and LAZZARINI, R. A. (1986). Matrix genes of measles virus and canine distemper virus: cloning, nucleotide sequence, and deduced amino acid sequences. *J. Virol.* **58,** 408–416.

BELLINI, W. J., ENGLUND, G., ROZENBLATT, S., ARNHEITER, H., and RICHARDSON, C. D. (1985). Measles virus P gene codes for two proteins. *J. Virol.* **53,** 908–919.

BUCKLAND, R., and WILD, F. (1989). Leucine zipper motif extends. *Nature (London)* **338,** 547.

CATTANEO, R., SCHMID, A., BILLETER, M. A., SHEPPARD, R. D., and UDEM, S. A. (1986). Multiple viral mutations rather than host factors cause defective measles virus gene expression in a subacute sclerosing panencephalitis cell line. *J. Virol.* **62,** 1388–1397.

CATTANEO, R., SCHMID, A., ESCHLE, D., BACZKO, K., TER MEULEN, V., and BILLETER, M. A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* **55,** 255–265.

CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J., and RUTTER, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **24,** 5294.

COSLETT, D. G., HOLLOWAY, B. P., and OBIJESKI, J. F. (1980). The structural proteins of rabies virus and evidence for their synthesis from separate monocistronic RNA species. *J. Gen. Virol.* **49,** 161–180.

COX, J. H., WEILAND, F., DIETZSCHOLD, B., and SCHNEIDER, L. G. (1981). Reevaluation of the structural proteins M1 and M2 of rabies virus. *In* "The Replication of Negative Strand Viruses" (D. H. L. Bishop and R. W. Compans, Eds.). Amer. Elsevier, New York.

DENTE, L., CESARENI, G., and CORTESE, R. (1983). pEMBL: A new family of single stranded plasmids. *Nucleic Acids Res.* **11,** 1645–1655.

DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387–395.

DIETZSCHOLD, B., LAFON, M., WANG, H., OTVOS, L., JR., CELIS, E., WUNNER, W. H., and KOPROWSKI, H. (1987). Localization and immunological characterization of antigenic domains of the rabies virus internal N and NS proteins. *Virus Res.* **8,** 103–125.

DIETZSCHOLD, B., WUNNER, W. H., WIKTOR, T. J., LOPEZ, A. D., LAFON, M., SMITH, M., and KOPROWSKI, H. (1983). Characterization of an antigenic determinant of the glycoprotein that correlates with pathogenicity. *Proc. Natl. Acad. Sci. USA* **80,** 70–74.

ERTL, H. C. J., DIETZSCHOLD, B., GORE, M., OTVOS, L., JR., LARSON, J. K., WUNNER, W. H., and KOPROWSKI, H. (1989). Induction of rabies virus-specific T-helper cells by synthetic peptides that carry dominant T-helper cell epitopes of the viral ribonucleoprotein. *J. Virol.* **63,** 2885–2892.

GALLIONE, C. J., and ROSE, J. K. (1983). Nucleotide sequence of a cDNA clone encoding the entire glycoprotein from the New Jersey serotype of vesicular stomatitis virus. *J. Virol.* **46,** 162–169.

GILL, D. S., and BANERJEE, A. K. (1985). Vesicular stomatitis virus NS proteins: Structural similarity without extensive sequence homology. *J. Virol.* **55,** 60–66.

GILMORE, R. D., JR., and LEONG, J.-A. C. (1988). The nucleocapsid gene of infectious hematopoietic necrosis virus, a fish rhabdovirus. *Virology* **167,** 644–648.

GIORGI, C., BLUMBERG, B. M., and KOLAKOFSKY, D. (1983). Sendai virus contains overlapping genes expressed from a single mRNA. *Cell* **35,** 829–836.

GUBLER, U., and HOFFMANN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25,** 263–269.

GUPTA, K. C., and KINGSBURY, D. W. (1984). Complete sequences of the intergenic and mRNA start signals in the Sendai virus genome: Homologies with the genome of vesicular stomatitis virus. *Nucleic Acids Res.* **12,** 3829–3841.

HENIKOFF, S. (1984). Unidirectional digestion of exonuclease III in DNA sequence analysis. *In* "Methods in Enzymology" (R. Wu, Ed.), Vol. 155, pp. 156–165. Academic Press, New York.

HERMAN, R. C. (1987). Vesicular stomatitis virus NS mRNA encodes two proteins. *In* "The Biology of Negative Strand Viruses" (B. Mahy and D. Kolakofsky, Eds.). Elsevier, Amsterdam/Oxford/New York.

HOLLOWAY, B. P., and OBIJESKI, J. F. (1980). Rabies virus-induced RNA synthesis in BHK21 cells. *J. Gen. Virol.* **49,** 181–195.

KOUZARIDES, T., and ZIFF, E. (1988). The role of the leucine zipper in the fos-jun interaction. *Nature (London)* **336,** 646–651.

KUCERA, P., DOLIVO, M., COULON, P., and FLAMAND, A. (1985). Pathways of the early propagation of virulent and avirulent rabies strains from the eye to the brain. *J. Virol.* **55,** 158–162.

KURILLA, M. G., CABRADILLA, C. D., HOLLOWAY, B. P., and KEENE, J. D. (1984). Nucleotide sequence and host La protein interactions of rabies virus leader RNA. *J. Virol.* **50,** 773–778.

LANDSCHULZ, W. H., JOHNSON, P. F., and MCKNIGHT, S. L. (1988). The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240,** 1759–1764.

LUK, D., SÁNCHEZ, Á., and BANERJEE, A. K. (1986). Messenger RNA encoding the phosphoprotein (P) gene of human parainfluenza virus 3 is bicistronic. *Virology* **153,** 318–325.

MASTERS, P. S., and BANERJEE, A. K. (1987). Sequences of chandipura virus N and NS genes: Evidence for high mutability of the NS gene within vesiculoviruses. *Virology* **157,** 298–306.

MCMASTER, G. K., and CARMICHAEL, G. G. (1977). Analysis of single and double stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74,** 4835–4838.

PEEPLES, M. E., and BRATT, M. A. (1984). Mutation in the matrix protein of newcastle disease virus can result in decreased fusion glycoprotein incorporation into particles and decreased infectivity. *J. Virol.* **51,** 81–90.

PREHAUD, C., COULON, P., DIALLO, A., MARTINET-EDELIST, C., and FLAMAND, A. (1989). Characterization of a new temperature-sensitive and avirulent mutant of the rabies virus. *J. Gen. Virol.* **70,** 133–143.

PREHAUD, C., COULON, P., LAFAY, F., THIERS, C., and FLAMAND, A. (1988). Antigenic site II of the rabies virus glycoprotein: Structure and role in viral virulence. *J. Virol.* **62,** 1–7.

RAYSSIGUIER, C., CIOE, L., WITHERS, E., WUNNER, W. H., and CURTIS, P. (1986). Cloning of rabies virus matrix protein mRNA and determination of its amino acid sequence. *Virus Res.* **5,** 177–190.

ROSE, J. K. (1980). Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. *Cell* **19,** 415–421.

ROSE, J. K., and GALLIONE, C. J. (1981). Nucleotide sequence of the mRNAs encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. *J. Virol.* **39,** 519–528.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

SCHNEIDER, L. G., and COX, J. H. (1983). Ein Feldversuch zur oralen Immunisierung von Füchsen gegen die Tollwut in der Bundesrepublik Deutschland. *Tierärztl. Umsch.* **5,** 315–328.

SCHNEIDER, L. G., COX, J. H., MÜLLER, W. W., and HOHNSBEEN, K.-P. (1988). Current oral rabies vaccination in Europe: an interim bilance. *Rev. Infect. Dis.* **10**(Suppl 4) S654–660.

SEIF, I., COULON, P., ROLLIN, P. E., and FLAMAND, A. (1985). Rabies virulence: Effect on pathogenicity and sequence characterization of rabies virus mutations affecting antigenic site III of the glycoprotein. *J. Virol.* **53,** 926–934.

TORDO, N., POCH, O., ERMINE, A., and KEITH, G. (1986a). Primary structure of leader RNA and nucleoprotein genes of the rabies genome: Segmented homology with VSV. *Nucleic Acids Res.* **14,** 2671–2683.

TORDO, N., POCH, O., ERMINE, A., KEITH, G., and ROUGEON, F. (1986b). Walking along the rabies genome: Is the large G-L intergenic region a remnant gene? *Proc. Natl. Acad. Sci. USA* **83,** 3914–3918.

TORDO, N., POCH, O., ERMINE, A., KEITH, G., and ROUGEON, F. (1988). Completion of the rabies virus genome sequence determination: Highly conserved domains among the L (Polymerase) proteins of unsegmented negative-strand RNA viruses. *Virology* **165,** 565–567.

WIKTOR, T. J., DIETZSCHOLD, B., LEAMNSON, R. N., and KOPROWSKI, H. (1977). Induction and biological properties of defective interfering particles of rabies virus. *J. Virol.* **21,** 626–635.

WUNNER, W. H., DIETZSCHOLD, B., SMITH, C. L., LAFON, M., and GOLUB, E. (1985). Antigenic variants of CVS rabies virus with altered glycosylation sites. *Virology* **140,** 1–12.

YELVERTON, E., NORTON, S., OBIJESKI, J. F., and GOEDDEL, D. V. (1983). Rabies virus glycoprotein analogs: Biosynthesis in Escherichia coli. *Science* **219,** 614–620.

ZIMMERN, D., and KAESBERG, P. (1978). 3′ terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **75,** 4257–4261.

ZHANG, H., SCHOLL, R., BROWSE, J., and SOMERVILLE, C. (1988). Double stranded DNA sequencing as a choice for DNA sequencing. *Nucleic Acids Res.* **16,** 1220.