VIROLOGY 168, 421–425 (1989)

Nucleotide Sequence of the Lassa Virus (Josiah Strain) S Genome RNA and Amino Acid Sequence Comparison of the N and GPC Proteins to Other Arenaviruses[1]

DAVID D. AUPERIN[2] AND JOSEPH B. MCCORMICK

Division of Viral Diseases, Center for Infectious Diseases, Centers for Disease Control, U.S. Public Health Serivce, Department of Health and Human Services, Atlanta, Georgia 30333

Received July 22, 1988; accepted October 12, 1988

The complete nucleotide sequence of the S genome RNA of the Josiah strain of Lassa virus was determined from cloned cDNA. The S RNA is 3402 nucleotides long with a calculated molecular weight of 1.09×10^6 Da. The nucleotide base composition is 26.84% adenine, 21.40% guanine, 22.75% cytosine, and 29.01% uridine. The 5′ and 3′ terminal nucleotide sequences are conserved and complimentary for 19 nucleotides, the nucleoprotein and glycoprotein genes are arranged in ambisense coding strategy, and the intergenic region contains an inverted complimentary sequence, as do all other arenavirus S RNAs characterized to date. Amino acid sequence comparisons between the nucleoproteins and glycoproteins of the Josiah and Nigerian (N sequences only) strains of Lassa virus, the WE and ARM strains of lymphocytic choriomeningitis virus (LCMV), Tacaribe, and Pichinde viruses are presented. These findings reveal that the G2 envelope glycoprotein is more conserved among different arenaviruses than the internal nucleoprotein.

Lassa virus, a member of the family Arenaviridae, is the etiologic agent of Lassa fever (*1, 2*). The disease is endemic in western and central Africa where the virus is maintained by persistent infection of the rodent species *Mastomys natalensis* (*3–5*). Upon infecting humans, which usually occurs via contact with persistently infected animals, the virus can cause a severe and often fatal hemorrhagic fever.

The Lassa virus genome consists of two pieces of single-stranded RNA designated large (L) and small (S) (*6, 7*). The S RNA encodes the viral nucleoprotein and glycoprotein genes in the ambisense organization characteristic of the Arenaviridae. A cDNA clone of the nucleoprotein gene of the Nigerian strain of Lassa virus has been constructed and expressed in vaccinia virus (*8*). Vaccination of guinea pigs with this recombinant resulted in protection from a lethal infection with Lassa virus. We have recently reported that a recombinant vaccinia virus expressing the glycoprotein gene of the Josiah strain of Lassa virus also protects guinea pigs (*9*) and primates (*10*) from a lethal Lassa virus infection.

To directly compare the protective efficacies of vaccinia virus recombinants expressing the nucleoprotein and glycoprotein genes, to provide cDNA clones of these genes for expression in other systems, and to provide sensitive hybridization probes for analyzing Lassa virus RNA for experimental and diagnostic purposes, we have completed the cloning and sequencing of the S genome RNA of the Josiah strain of Lassa virus. We report the results of this work here, including amino acid sequence comparisons of the N and GPC proteins of Lassa virus to those of lymphocytic choriomeningitis (LCM), Tacaribe, and Pichinde viruses.

Three preparations of cDNA were synthesized from viral RNA templates and blunt-end ligated into the unique *Sma*I site of pUC18 for sequence analysis. Initially, a DNA oligonucleotide complimentary to the 19-nucleotide conserved sequence on the 3′ end of Arenavirus S RNAs was used to prime cDNA synthesis. This produced a library from which two overlapping clones, LS55 and LS13, corresponding to nucleotides 620–1974 and 1745–2454, respectively, were isolated. To protect and conserve the 3′ terminus of the S RNA, poly(A) tails were enzymatically added and a second cDNA preparation was made by oligo(dT)-primed first-strand synthesis. Clone LS60-7 was isolated from the resulting library, and sequence analysis revealed it to contain 30 A residues reading into the conserved sequence on the 3′ end of the S RNA. This clone terminated at position 1032. Finally, a cDNA clone extending to the 5′ end of the Lassa S RNA was produced by priming first-strand cDNA synthesis with the viral complimentary strand of a *Dde*I restriction fragment corresponding to positions 2110–2410. The single-strand cDNA product of this reaction was resolved by polyacrylamide gel electrophoresis, its 5′ end was protected by terminal transferse tailing, and it was subsequently converted to dsDNA and cloned. Clone

[1] Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. J04324.

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

0042-6822/89 $3.00

```
      10        20        30        40        50        60        70        80        90       100       110       120
CGCACAGUGGAUCCUAGGCUAUUGGAUUGCGCUUUGCUUUUGUCAUUUUGGCAGAUAGUCUCAGUUCUUUGUUGCGUGCAUACAACACAACAAUCUGGCGAUGAGUGCCUCAAAGGAAAU
                                                                                                    METSerAlaSerLysGluIle

     130       140       150       160       170       180       190       200       210       220       230       240
AAAAUCCUUUUUGUGGACACAAUCUUUGAGGAGGGAAUUAUCUGGUUACUGCUCCAACAUCAAACUACAGGUGGUGAAAGAUGCCCAGGCUCUUUUACAUGGACUUGACUUCUCCGAAGU
 LysSerPheLeuTrpThrGlnSerLeuArgArgGluLeuSerGlyTyrCysSerAsnIleLysLeuGlnValValLysAspAlaGlnAlaLeuLeuHisGlyLeuAspPheSerGluVal

     250       260       270       280       290       300       310       320       330       340       350       360
CAGUAAUGUUCAACGGUUGAUGCGCAAGGAGAGAAGGGAUGACAAUGAUUUGAAACGGUUGAGGGACCUAAAUCAAGCGGUCAACAAUCUUGUUGAAUUAAAAUCAACUCAACAAAAGAG
 SerAsnValGlnArgLeuMETArgLysGluArgArgAspAspAsnAspLeuLysArgLeuArgAspLeuAsnGlnAlaValAsnAsnLeuValGluLeuLysSerThrGlnGlnLysSer

     370       380       390       400       410       420       430       440       450       460       470       480
UAUACUGAGAGUUGGGACUCUAACCUCAGAUGACUUAUUAAUCUUAGCCGCUGAUCUAGAGAAGUUAAAGUCAAAGGUGAUCAGAACAGAAAGGCCAUUAAGUGCAGGUGUCUAUAUGGG
 IleLeuArgValGlyThrLeuThrSerAspAspLeuLeuIleLeuAlaAlaAspLeuGluLysLeuLysSerLysValIleArgThrGluArgProLeuSerAlaGlyValTyrMETGly

     490       500       510       520       530       540       550       560       570       580       590       600
CAACCUAAGCUCACAGCAACUUGACCAAAGAAGAGCUCUCCUGAAUAUGAUAGGAAUGAGUGGUGGUAAUCAAGGGGCUCGGGCUGGGAGAGAUGGAGUGGUGAGAGUUUGGGAUGUGAA
 AsnLeuSerSerGlnGlnLeuAspGlnArgArgAlaLeuLeuAsnMETIleGlyMETSerGlyGlyAsnGlnGlyAlaArgAlaGlyArgAspGlyValValArgValTrpAspValLys

     610       620       630       640       650       660       670       680       690       700       710       720
AAAUGCAGAGUUGCUCAAUAAUCAGUUCGGGACCAUGCCAAGUCUGACACUGGCAUGUCUGACAAAACAGGGGCAGGUUGACUUGAAUGAUGCAGUACAAGCAUUGACAGAUUUGGGUUU
 AsnAlaGluLeuLeuAsnAsnGlnPheGlyThrMETProSerLeuThrLeuAlaCysLeuThrLysGlnGlyGlnValAspLeuAsnAspAlaValGlnAlaLeuThrAspLeuGlyLeu

     730       740       750       760       770       780       790       800       810       820       830       840
GAUCUACACAGCAAAGUAUCCCAACACUUCAGACUUAGACAGGCUGACUCAAAGUCAUCCCAUCCUAAAUAUGAUUGACACCAAGAAAAGCUCUUUGAAUAUCUCAGGUUAUAAUUUUAG
 IleTyrThrAlaLysTyrProAsnThrSerAspLeuAspArgLeuThrGlnSerHisProIleLeuAsnMETIleAspThrLysLysSerSerLeuAsnIleSerGlyTyrAsnPheSer

     850       860       870       880       890       900       910       920       930       940       950       960
CUUGGGUGCAGCUGUGAAGGCAGGAGCUUGCAUGCUGGAUGGUGGCAAUAUGUUGGAGACAAUCAAGGUGUCACCUCAGACAAUGGAUGGUAUCCUCAAAUCCAUUUUAAAGGUCAAGAA
 LeuGlyAlaAlaValLysAlaGlyAlaCysMETLeuAspGlyGlyAsnMETLeuGluThrIleLysValSerProGlnThrMETAspGlyIleLeuLysSerIleLeuLysValLysLys

     970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
GGCUCUUGGAAUGUUCAUUUCAGACACCCCUGGUGAAAGGAAUCCUUAUGAAAACAUACUCUACAAGAUUUGUUUGUCAGGAGAUGGAUGGCCAUAUAUUGCAUCAAGAACCUCAAUAAC
 AlaLeuGlyMETPheIleSerAspThrProGlyGluArgAsnProTyrGluAsnIleLeuTyrLysIleCysLeuSerGlyAspGlyTrpProTyrIleAlaSerArgThrSerIleThr

     1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
AGGAAGGGCCUGGGAAAACACUGUCGUUGAUCUGGAAUCAGAUGGGAAGCCACAGAAAGCUGACAGCAACAAUUCCAGUAAAUCCCUGCAGUCGGCAGGGUUUACCGCUGGGCUUACCUA
 GlyArgAlaTrpGluAsnThrValValAspLeuGluSerAspGlyLysProGlnLysAlaAspSerAsnAsnSerSerLysSerLeuGlnSerAlaGlyPheThrAlaGlyLeuThrTyr

     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
UUCUCAGCUGAUGACCCUCAAGGAUGCAAUGCUGCAACUUGACCCAAAUGCUAAGACCUGGAUGGACAUUGAAGGAAGACCUGAAGAUCCAGUGGAAAUUGCCCUCUAUCAACCAAGUUC
 SerGlnLeuMETThrLeuLysAspAlaMETLeuGlnLeuAspProAsnAlaLysThrTrpMETAspIleGluGlyArgProGluAspProValGluIleAlaLeuTyrGlnProSerSer

     1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440
AGGCUGCUACAUACACUUCUUCCGUGAACCUACUGAUUUAAAGCAGUUCAAGCAGGAUGCUAAGUACUCACAUGGGAUUGAUGUCACAGACCUCUUCGCUACACAACCGGGCUUGACCAG
 GlyCysTyrIleHisPhePheArgGluProThrAspLeuLysGlnPheLysGlnAspAlaLysTyrSerHisGlyIleAspValThrAspLeuPheAlaThrGlnProGlyLeuThrSer

     1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560
UGCUGUCAUUGAUGCACUCCCCCGGAAUAUGGUCAUUACCUGUCAGGGGUCCGAUGACAUAAGGAAACUCCUUGAAUCACAAGGAAGAAAAGACAUUAAACUAAUUGAUAUUGCCCUCAG
 AlaValIleAspAlaLeuProArgAsnMETValIleThrCysGlnGlySerAspAspIleArgLysLeuLeuGluSerGlnGlyArgLysAspIleLysLeuIleAspIleAlaLeuSer

     1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680
CAAAACUGAUUCCAGGAAGUAUGAAAAUGCAGUCUGGGACCAGUAUAAAGACUUAUGCCACAUGCACACAGGUGUCGUUGUUGAAAAGAAGAAAAGAGGCGGUAAAGAGGAAAUAAACCC
 LysThrAspSerArgLysTyrGluAsnAlaValTrpAspGlnTyrLysAspLeuCysHisMETHisThrGlyValValValGluLysLysLysArgGlyGlyLysGluGluIleThrPro

     1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
UCACUGUGCACUAAUGGACUGCAUCAUGUUUGAUGCAGCAGUGUCAGGAGGACUGAACACAUCGGUUUUGAGAGCAGUGCUGCCCAGAGAUAUGGUGUUCAGAACAUCGACACCUAGAGU
 HisCysAlaLeuMETAspCysIleMETPheAspAlaAlaValSerGlyGlyLeuAsnThrSerValLeuArgAlaValLeuProArgAspMETValPheArgThrSerThrProArgVal

     1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920
CGUUCUGUAAAUGGACGCCCCCGUGACCCACCGCCAAUAGGCGGUGGGUCACGGGGGCCCUGACAAGGGUCUCAUCUCUUCCAUUUCACAGGCACACCAGGCUGUUUGUAGAGUCCACAG
 ValLeuend

     1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040
GAACAAAUGCCCAUAUGAUUCAAUCUGUGAGGUUUGGGACACGACUUGCCUACAAUAUGCCUAUGAGUUGGUAUUUUGACUAGGUGAAGGAAGAUGCUAAUAAGAUAGAAACUUGUACUG

     2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160
AACACAAAGAGGUCAACUAGACCCAAUGGUGUCUUCCCCUGCCUCUCCAUAUACUCCUUCUGUAACAUCUCAGUGAUCAUAUUGUCAGCUUGUUGUUCAAUAUCAUCAGAAAAGUGGGUC

     2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280
UCGUUCAAGUAUGAACCAUUUGAUACAAGCCAACAUUUGGGCAGUGAUGUUCUCCCAGUAGUUGUGUGGUUGAGGUACCAAUACUUGCUGUAAUUACAGUAUGGAAUUCCCAUGAUGUCC

     2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
CGUAGAUGGUUCUUCAUUAUAAGUUGGUCAUUUAUCAAAGCAUUUACUGCUUUGUUGAUCAACUGAAUGCUCAUUUGUGCUUCAGCUUUCAACCUUUGAAUGGCUUGUUUGUUGAAGUCA

     2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
AACAGCCUCAGCAUGUCACAAAAUUCCUCAUCAUGCUUCUCAUUACAUUUUGCCACAGCUGUGUUCCCGAAGCAUUUUAGUUCAGCCUCAAUUAGCAUCCACCUGGUCAGACAAAUAUCCC

     2530      2540      2550      2560      2570      2580      2590      2600      2610      2620      2630      2640
CCUGGUGUGUCUUUACCUUCAGAAUCUGACAGUGUCCAUGUGAAUGUGCCUAGCAAUCUUCUACUAAUAUAAAUAUCUCUAGUCCUUUGUGAGAGGAGCCCGAGAUAACCGAUGGGAGAU

     2650      2660      2670      2680      2690      2700      2710      2720      2730      2740      2750      2760
GGUCUCGAGAAUUGGCAGUGAUCUUCCCAGGUUGUAUUUUGGAUUAUCAGAUAUUGAUAACUAGUCAUAAUACAGUCCCAGUUGCCACGGCCUGAGUCAAGAGCAAUGUAGCUCCCACCC

     2770      2780      2790      2800      2810      2820      2830      2840      2850      2860      2870      2880
CAAGCCAUCCUCAUAAAAGUCUGUAACACACCAUUUGCAACAGUACCACAAUGGUUUGGCUGCAUCCCCAGCAUAGCUGUGACUCAGGUUGUACUGCACACUAAUCUUUCCCCCAUUAAAA

     2890      2900      2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
UCGCAGCUCAUUGCCUCAUACUGAUUGAAGUUGGGGAUGGACAAGUGGAAAGUUGAGAUUAUGCUCAUAAGAGCGUGGUCAUAGAGGUUCUUUUUGUGGGCAUCAGACAGAUUGCAAAAU

     3010      3020      3030      3040      3050      3060      3070      3080      3090      3100      3110      3120
UUGUGAUUAAUAAUGCUCGUGUUGGUCAAGGUCAGUUCUAGUCCUGUCUCAUUGCCCACCAUUAUAUAAUGAUGACUGUUGUUCUUUGUGCAGGAGAGAGGCAUGGUCAUAUUGAGUGUC

     3130      3140      3150      3160      3170      3180      3190      3200      3210      3220      3230      3240
UCCAUGUUUAGUUCCAGAGUCUGAAGCUCAUAAACCCCUUUAUAAAGACUGGUUGUGCAAGACCUACCACACAACAGGAGGAAAGUGACCAAACCAACAAGGCCACACGUUGCAAAAUUG

     3250      3260      3270      3280      3290      3300      3310      3320      3330      3340      3350      3360
UACAGACCUUUCAGCACUGCUAGUACAGACAGUGCAAUGAGAACAAUGUUCAUCACCUCUUCUAUUACAUGAGGCACUUCCUGGAAGAAUGUCACUAUUUGUCCCAUUUUAAAUAGGACA

     3370      3380      3390      3400
CUUGAATTGCGCAACCAAAAAUGCCUAGGAUCCCCGGUGCGC
```

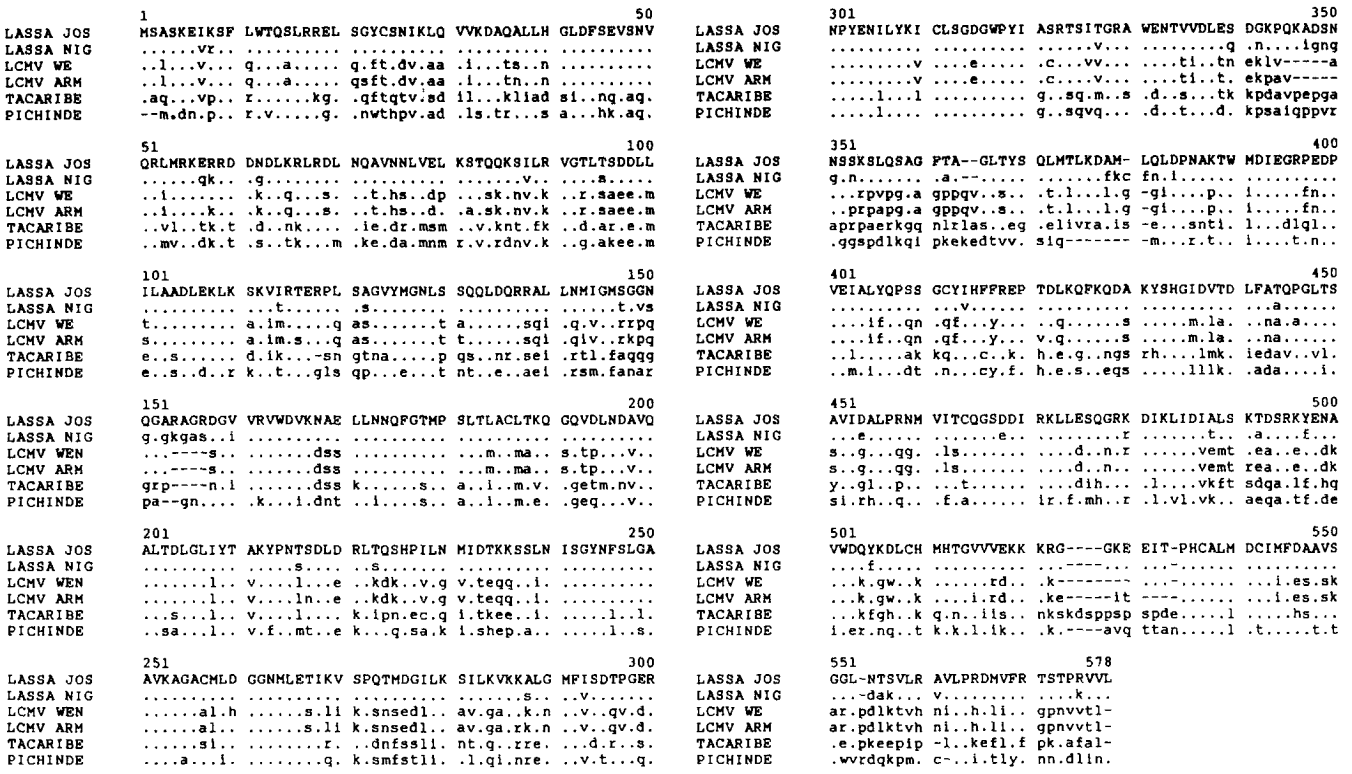

FIG. 2. Amino acid sequence comparisons of the nucleoproteins of the Josiah and Nigerian strains of Lassa virus, the WE and Armstrong strains of LCMV, Tacaribe, and Pichinde viruses. The alignments were generated by reiterative comparisons between two amino acid sequences until no additional gaps were inserted using the algorithm of Needleman and Wunsch (*16*) as supplied in the University of Wisconsin Genetics Computer Group software package (*17*). The Lassa Josiah nucleoprotein was used as the reference sequence, and the alignments were made without allowing the substitution of evolutionarily related amino acids within a given sequence. Dots (···) indicate identity with the reference sequence and dashes (- - -) denote gaps introduced by the alignment algorithm.

LS37 was isolated from this library and found to originate precisely at position 2110 and terminate at position 3402 with a C residue reading into a homopolymer tail of 22 A nucleotides.

Figure 1 presents the complete nucleotide sequence of the Lassa virus S RNA given as viral complimentary RNA and the amino acid sequence of the nucleoprotein it encodes. The S RNA is 3402 nucleotides long with a calculated molecular weight of 1.09×10^6 Da. The nucleotide base composition is 26.84% A, 21.40% G, 22.75% C, and 29.01% U. The 3′ and 5′ termini of the Lassa virus S RNA are conserved and complimentary for 19 nucleotides, as has been observed for other arenaviruses. This sequence presumably represents a binding site for the viral-specific RNA polymerase to initiate RNA synthesis.

The nucleoprotein coding sequence initiates at positions 101–103 and terminates with a UAA stop codon at positions 1808–1810. The protein contains 569 amino acids with a calculated molecular weight of 63,006 Da and a net charge of +9 at neutral pH. The glycoprotein coding sequence initiates at positions 3347–3345 and terminates with a UGA stop codon at positions 1874–1872 in the viral RNA sequence. The intergenic region located between the termination codons of these two genes contains an 18-nucleotide inverted complimentary sequence (1818–1857) capable of forming a hairpin structure similar to the S RNAs of other arenaviruses. The conservation of such a structure by all members of the Arenavirus family sequenced to date (Lassa, LCMV, Pichinde, Tacaribe) suggests its importance in regulating RNA replication and gene transcription during productive as well as persistant infection.

FIG. 1. Viral complimentary RNA nucleotide sequence of the Lassa virus (Josiah strain) S genome RNA. The amino acid sequence of the nucleoprotein is given beneath the nucleotide codons. Overlined sequence corresponds to the inverted complimentary sequence within the intergenic region. Underlined sequence delineates the glycoprotein gene encoded by viral RNA sequence. Nucleotide sequences were determined by the method of Maxam and Gilbert (*11*) on cloned cDNA. Double-strand data were obtained for greater than 95% of the sequence.

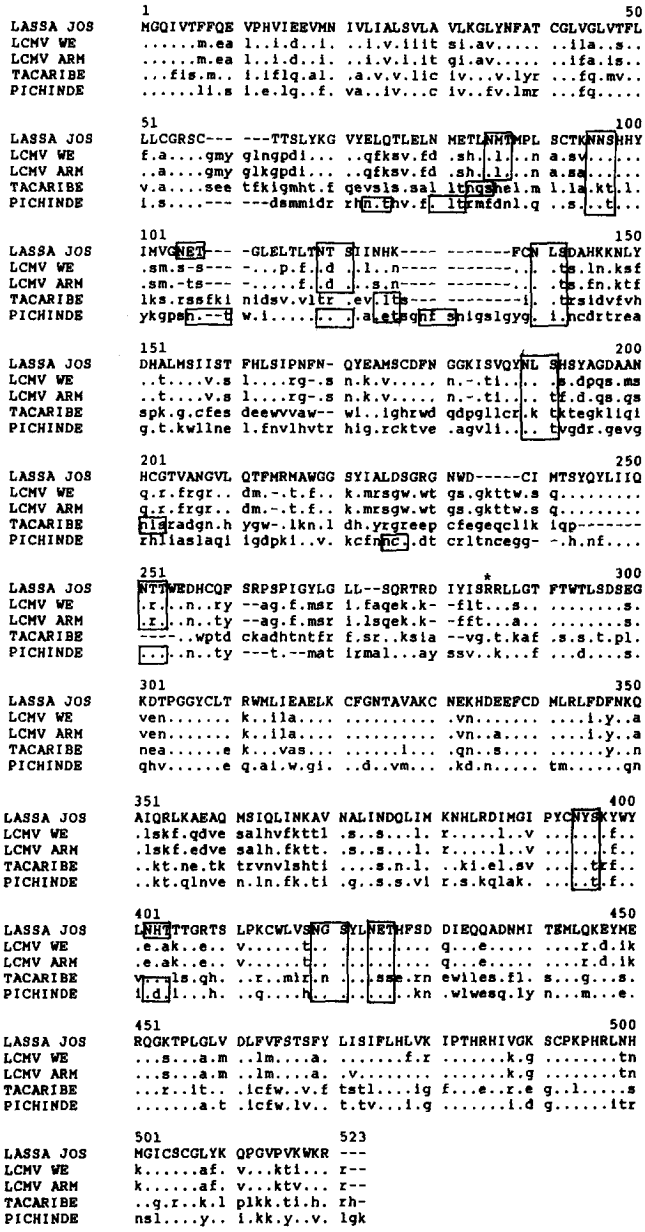

FIG. 3. Amino acid sequence comparison of the glycoprotein precursors of Lassa virus (Josiah strain), WE and Armstrong strains of LCMV, Tacaribe, and Pichinde viruses. The alignments were performed as described in Fig. 2. The potential asparagine-linked glycosylation sites are boxed. The (*) denotes the position of the proteolytic cleavage site (ArgArg or ArgLys) between the amino-terminal G1 protein and the carboxy-terminal G2 protein of Lassa, LCMV, and Pichinde viruses.

Figure 2 shows a computer-generated alignment of the amino acid sequences of the nucleoproteins of the Josiah and Nigerian strains of Lassa virus, the WE and ARM strains of LCMV, Tacaribe, and Pichinde viruses (*6, 7, 12–15*). The N proteins of the two Lassa virus and the two LCMV strains are very similar to each other, with 90.86 and 95.52% amino acid sequence similarity, respectively. The percentage of amino acid sequence similarity between the N proteins of Lassa (JOS) and LCMV (WE) is 63.78%, between Las (JOS) and Tacaribe is 50.44%, and between Tacaribe and Pichinde is 56.73%. The most conserved regions of the protein correspond to amino acid residues 237–268 and 300–338. Four cystein residues at positions 186, 311, 509, and 537 are conserved by all six viruses.

Figure 3 shows a similar alignment of the amino acid sequences of the glycoprotein precursor of the Josiah strain of Lassa virus, the WE and ARM strains of LCMV, Tacaribe, and Pichinde viruses. The G1 portion of the molecule corresponds to amino acids 1–285 in the aligned sequence, where a dibasic amino acid residue occurs that was previously shown to be the cleavage site for LCMV (*18*). This site is conserved by all viruses except Tacaribe, which contains an ArgThr instead of an ArgArg or an ArgLys. Since this amino acid substitution cannot be accounted for by a single base mutation, and all evidence to date indicates that Tacaribe virus possesses only one glycoprotein of approximately 42,000 Da (*19*); the details of its synthesis and processing remain obscure. The G2 portion of the molecule corresponds to amino acid residues 286–523. It is apparent from this alignment that the G2 proteins are considerably more conserved among individual members of the Arenaviridae than the G1 proteins. There are three asparagine-linked glycosylation sites between amino acid residues 394 and 424 that are conserved by all six viruses. In addition, Lassa (JOS), Tacaribe, and Pichinde share a fourth glycosylation site at position 402. The G1 proteins of arenaviruses have the

TABLE 1

PERCENTAGE AMINO ACID SEQUENCE SIMILARITY BETWEEN THE N, G1, AND G2 PROTEINS OF LASSA, LCM, TACARIBE, AND PICHINDE VIRUSES

Viruses compared	Proteins		
	N	G1	G2
Las JOS vs LCMV WE	63.78	50.97	70.21
Las JOS vs LCMV ARM	62.29	52.99	71.06
LCMV ARM vs LCMV WE	95.52	91.99	96.19
Las JOS vs Pichinde	50.54	36.76	57.87
LCMV WE vs Pichinde	50.00	31.62	54.66
LCMV ARM vs Pichinde	50.00	31.47	54.24
Tacaribe vs Las JOS	50.44	24.19	53.19
Tacaribe vs LCMV WE	50.54	23.32	51.70
Tacaribe vs LCMV ARM	51.62	20.16	51.66
Tacaribe vs Pichinde	56.73	32.00	54.01

Note. Amino acid sequence comparisons were performed as described in Fig. 2. The G1 and G2 proteins were defined by the cleavage site identified for LCMV at position 285 in the aligned sequence. For this comparison, the G1 and G2 portions of the Tacaribe GPC were defined similarly even though two glycoproteins have not been observed for this virus.

potential for more glycosylation, specifically between amino acid residues 73 and 140, where depending upon the virus are four (LCM ARM) to eight (Pichinde) glycosylation sites. Three glycosylation sites at positions 95, 140, and 189 are conserved by all six viruses.

Table 1 compares the percentage of amino acid similarity (identity and position) between the N, G1, and G2 proteins of Lassa Josiah, LCM WE and Armstrong, Tacaribe, and Pichinde viruses. These findings reveal that the G2 protein is slightly more conserved by these viruses than the N protein in all but one comparison (Tacaribe vs Pichinde). The significance of this is not understood, but it may suggest that the G2 envelope glycoprotein provides a conserved function at some point in the infectious cycle, possibly during morphogenesis or attachment and entry into susceptible cells. As expected, the more variable G1 proteins exhibit the least amino acid sequence similarity.

Now that the genes for the internal nucleoprotein and envelope glycoproteins of Lassa virus have been cloned, sequenced, and assembled for expression, their potential for developing diagnostic antigens, nucleic acid hybridization probes, and live recombinant virus vaccines can be explored. In addition, expression of the Lassa virus N and GPC genes in tissue culture can now be applied to the study of gene function in arenavirus replication.

ACKNOWLEDGMENT

The authors are grateful for the excellent technical support provided by Donna R. Sasso in the isolation and sequence analysis of the cDNA clones.

REFERENCES

1. Frame, J. D., Baldwin, J. M., Jr., Gocke, D. J., and Troup, J. M., *Amer. J. Trop. Med. Hyg.* **19,** 670–676 (1970).
2. Buckley, S. M., and Casals, J., *Amer. J. Trop. Med. Hyg.* **19,** 680–691 (1970).
3. Monath, T. P., Newhouse, V. F., Kemp, G. E., Setzer, H. W., and Cacciapuoti, A., *Science* **185,** 262–265 (1974).
4. Murphy, F. A., and Walker, D. H., *In "Viruses and Environment"* (E. Kurstak and K. Maramorosch, Eds.), pp. 155–180. Academic Press, New York, 1978.
5. Walker, D. H., Wulff, H., Lange, J. V., and Murphy, F. A., *Bull. W.H.O.* **52,** 523–534 (1975).
6. Clegg, J. C. S., and Oram, J. D., *Virology* **144,** 363–372 (1985).
7. Auperin, D. D., Sasso, D. R., and McCormick, J. B. M., *Virology* **154,** 155–167 (1986).
8. Clegg, J. C. S., and Lloyd, G., *Lancet* **2,** 186–188 (1987).
9. Auperin, D. D., Esposito, J. J., Lange, J. V., Bauer, S. P., Knight, J., Sasso, D. R., and McCormick, J. B. M., *Virus Res.* **9,** 233–248 (1988).
10. Fisher-Hoch, S. P., McCormick, J. B., Auperin, D. D., Brown, B. G., Castor, M., Perez, G., Ruo, S., Conaty, A., Bramer, L., and Bauer, S., *Proc. Natl. Acad. Sci. USA,* in press.
11. Maxam, A., and Gilbert, W., *In "Methods in Enzymology"* (L. Grossman and K. Moldave, Eds.), Vol. 65, pp. 499–560. Academic Press, New York, 1980.
12. Romanowski, V., Matsuura, Y., and Bishop, D. H. L., *Virus Res.* **3,** 101–114 (1985).
13. Southern, P. J., Singh, M. K., Riviere, Y., Jacoby, D. R., Buchmeier, M. J., and Oldstone, M. B. A., *Virology* **157,** 145–155 (1987).
14. Franze-Fernandez, M-T., Zetina, C., Iapalucci, S., Lucero, M. A., Bouissou, C., Lopez, R., Rey, O., Daheli, M., Cohen, G. N., and Zakin, M., *Virus Res.* **7,** 309–324 (1987).
15. Auperin, D. D., Romanowski, V., Galinski, M., and Bishop, D. H. L., *J. Virol.* **52,** 897–904 (1984).
16. Needleman, A., and Wunsch, C. D., *J. Mol. Biol.* **48,** 443–453 (1970).
17. Devereux, J., Haeberli, P., and Smithies, O., *Nucleic Acids Res.* **12,** 387–395 (1984).
18. Buchmeier, M. J., Southern, P. J., Parekh, B. S., Wooddell, M. K., and Oldstone, M. B. A., *J. Virol.* **61,** 982–985 (1987).
19. Saleh, F., Gard, G. P., and Compans, R. W., *Virology* **93,** 369–376 (1979).