

**Nucleocapsid protein gene of Junin arenavirus (cDNA sequence)**

P.D.Ghiringhelli, R.V.Rivera Pomar, N.I.Baro, M.F.Rosas, O.Grau and V.Romanowski\*

Area de Química Biológica y Biología Molecular, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, calle 47 y 115, 1900 La Plata, Argentina  
 Submitted August 9, 1989

EMBL accession no. X15827

The Arenavirus S RNA codes for the major structural polypeptides: the nucleocapsid protein, N, and the glycoproteins G1 and G2, that are derived from an intracellular precursor protein, GPC.

We have cloned and sequenced the entire coding region for the N gene of Junin arenavirus, the etiologic agent of Argentine hemorrhagic fever. The cDNA sequence was determined by a dideoxy procedure (1) using a modified T7 DNA polymerase (Sequenase™, USB, Cleveland, Ohio, USA) and is shown as the viral-complementary strand.

The upstream non-coding sequence comprises 81 nucleotides, and the open reading frame codes for the N polypeptide, deduced to be 564 aminoacids long (mol. wt. 63,033 Da). Nucleotide and aminoacid sequences exhibit varying degrees of identity when compared to homologous regions of other arenavirus sequences. The comparison with Tacaribe virus (2), an arenavirus with a close antigenic relation to Junin, shows a 76% identity in aminoacid sequence, whereas with N proteins of other arenaviruses the identity ranges from 46 to 52%.

```

M A H S K E V P S F R W T 13
CGCACAGTGGATCTCCAGCTAGTGTCTAACACTTTCACGTGAGCAATTTTCTCAGATCAAACCACCAGATCTCTGGCATGGACACTCAAAGAGGTTCCAGCTTTAGATGGACT 120
G L S Q F T O T V K S D V L K Q S L R R D A K L I A D S I D F N Q V A Q V Q R A 53
GGTTTGAGTCAATTTACTCAGACTGTCAAGTCGGATGTTCTGAAGCAGTCCCTTAAGGAGAGATGCCAAGCTAATTTGCTGACAGCATTGACTCAACCAAGTGGCACAGGTGCAGCGGGCA 240
L R K T K R G E E D L N K L R D L N K E V D R L M S M R S V Q R N T V F K A G D 93
CTCAGAAAGACTAAAAGGGGGGAAGAAGACTCAATAGTGTGAGGACCTGAATAAGAGGTTGACAGACTCATGTCCATGAGGAGTGTTCACGAACCAAGCGGTTTTCAAGCGGGGAGAT 360
L G R V E R M E L A S G L G N L K T K F R R A E T G S Q G V Y H G N L S Q S Q L 133
CTGGGGAGGGTTGAGCGGATGGAGTTGGCGTCTGGCCTTGGGAATTTAAAACCAAGTTTGAAGAGCAGAGACAGGCTCAAGGGGTTTACATGGGTAACTGTCTCAGTCAACAATG 480
A K R S E I L R T L G F Q Q Q G T G G H G V V R V M D V K D P S K L N H Q F G S 173
GCCAAGAGATCAGAGATATGAGAACCTGGGATTTCAACAGCAAGAACTGGGGGAATGGTGGTGGTGGGATGGGATGTTAAAGACTCTTCAAACCTTAAACAATCAGTTTGGTCT 600
V P A L T I A C M T V Q G G E T M N S V I Q A L T S L G L L Y T V K Y P H L S D 213
GTCTCGCATGACAATTTGATGCATGCTTCAAGGAGGTGAGACAATGAACAGTGTCTACAGGCTTTAACTCTCTTGGGCTTATACACTGTGAGATGTCACCAACTTGAGTGAC 720
L D R L T Q E H D C L Q I V T K D E S S I N I S G Y N F S L S A A V K A G A S I 253
CTTGACAGATTGACTCAGGAACATGACTGCCGTCAGATTTGACTAAGGATGAAGCTCCATCAACATTTCTGGCTACAACCTTTAGTCTTTCCAGTGCAGTAAAGGCTGGTGCATCATT 840
L D D G N H L E T I R V T P D N F S S L I K S T I Q V K R R E G H F I D E K P G 293
CTTGATGATGAAATATGTTGGAGACAATCAGAGTCAACCCAGATAACTTTTCTCCCTCAAAAATCAACCTTCAAGGTTAAACGAAGAGAGGCACTGTTATTGATGAGAAACCAAGC 960
H R N P Y E N L L Y K L C L S G D G W P Y I G S R S Q I I G R S W D N T S I D L 333
AATGAAACCTTTATGAAACCTTCTGACAACTTTGCCCTTCTGGCGATGGTGGCCCTATATTGGTCCAGATCACAATCATAGGCAGGTCATGGGACCAACCAAGATGATGATCTG 1080
T R K P V A G P R Q P E K N G O N L R L A N L T E I Q E A V I R E A V G K L D P 373
ACAAGGAACCAAGTCTGCTGACCTAGACAACCGGAAGAAGCGGTGAGAATTTGAGATTGGCTAACTTGACAGAGATACAAGAAGCTGTATCAGAGAGGCACTGGGGAACCTGCACCCC 1200
T H T L W L D I E G P A T D P V E H A L F Q P A G S K Y I N C F R K P H D E K G 413
ACCAATACCCCTTGGCTCGACATGAAGGACCAGCTACTGACCTGTTGATGGCGTGTTCAGCCTGCAGGTAGCAAGTATTCCTGCTTCAAGAAACCAACATGATGAGAAGGG 1320
F K N G S R H S H G I L M K D I E D A N P G V L S Y V I G L L P P D H V Y T T Q 453
TTTAAAAATGGTAGGCACACTCTCAGGCTCTTAATGAAGACATAGAAGTGAATGCCAGGATCTTAGTTACGTGATGGCTGCTGCCCTCCGACATGGTGTGACCACTCAA 1440
G S D D I R K L F D L H G R R D L K L V D V R L T S E Q A R Q F D Q Q V W E K F 493
GGTCCGATGACATCAGGAAGTGTGTTGACCTCCATGAAGAAGAGATCTTAAGCTGGTGTAGTGTAGGCTTACATCAGAACAAGCCAGGCACTGACCAACAGGCTCTGGGAGAAATTT 1560
G H L C K H N G V V S K K R D K D A P F K L A S S E P H C A A L L D C I N F 533
GGCCACCTATGCAAAACATCAAAATGGAGTGGTGTGACGAAGAAGAAGAGATAGGATGCTCCCTTTAAGTGGCCTCCAGTGAGCCACACTGTCTCTCTAGACTGCATATGTTT 1680
Q S V L D G K L Y E E E L T P L L P S L L F L P K A A Y A L * 564
CAGTCAGTGTGGATGGAGACTCTATGAAGAAGAACTTACACTCTATTACCACCGAGCTTGTCTTCTCCCGAAGGACGCTTATGCATGTAAGCAGTGCCTCCCGACTCCCGGCC 1800
    
```

**Acknowledgements:**

This work has been supported by grants from CONICET, SECYT and CIC BA (Argentina).

\*To whom correspondence should be addressed

**References:**

1. Sanger *et al.* (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5466.
2. Franze-Fernandez *et al.* (1987) *Virus Res.* 7, 309-324.