Complete sequence of the genome of the human isolate of Andes virus CHI-7913: comparative sequence and protein structure analysis

NICOLE D. TISCHLER,[1,2] JORGE FERNÁNDEZ,[6] ILSE MÜLLER,[5] RODRIGO MARTÍNEZ,[1,2] HÉCTOR GALENO,[6] ELIECER VILLAGRA,[6] JUDITH MORA,[6] EUGENIO RAMÍREZ,[6] MARIO ROSEMBLATT,[1,2,3] and PABLO D.T. VALENZUELA[1,2,3,4]

Fundación Ciencia para la Vida[1], Instituto Milenio MIFAB[2], Universidad Andrés Bello[3], Pontificia Universidad Católica de Chile[4], Bios Chile IGSA[5]. Av. Marathon 1943, Santiago, Chile, Instituto de Salud Pública de Chile[6]. Av. Marathon 1000, Santiago, Chile

ABSTRACT

We report here the complete genomic sequence of the Chilean human isolate of Andes virus CHI-7913. The S, M, and L genome segment sequences of this isolate are 1,802, 3,641 and 6,466 bases in length, with an overall GC content of 38.7%. These genome segments code for a nucleocapsid protein of 428 amino acids, a glycoprotein precursor protein of 1,138 amino acids and a RNA-dependent RNA polymerase of 2,152 amino acids. In addition, the genome also has other ORFs coding for putative proteins of 34 to 103 amino acids. The encoded proteins have greater than 98% overall similarity with the proteins of Andes virus isolates AH-1 and Chile R123. Among other sequenced Hantavirus, CHI-7913 is more closely related to Sin Nombre virus, with an overall protein similarity of 92%. The characteristics of the encoded proteins of this isolate, such as hydrophobic domains, glycosylation sites, and conserved amino acid motifs shared with other Hantavirus and other members of the *Bunyaviridae* family, are identified and discussed.

Key terms: Hantavirus, genome sequence, RNA polymerase, human isolate

INTRODUCTION

Hantaviruses are enveloped rodent-borne negative sense RNA viruses that cause either hemorrhagic fever with renal syndrome or Hantavirus pulmonary syndrome (HPS), a severe and often fatal disease in humans (Schmaljohn and Hjelle, 1997).

Like other viruses in the *Bunyaviridae* family, Hantaviruses posses a tripartite RNA genome. The small or S RNA segment encodes the nucleocapsid protein N, the medium or M RNA segment encodes a glycoprotein precursor that is cleaved into the envelope glycoproteins G1 and G2, and the large or L RNA segment, which encodes a RNA-dependent RNA polymerase.

Andes virus (ANDV) belongs to the genus Hantavirus, which was first reported in 1995 to cause several fatal cases of HPS in Argentina (López *et al.*, 1996) and later recognized to be genetically and serologically different from others in the world and responsible for most of the HPS cases in Argentina and Chile (López *et al.*, 1997). In Chile, 276 cases of HPS have been reported through November 2002 with a 40% mortality rate.

Up to know, knowledge of the genetic organization of ANDV has been limited to the sequence of the isolate Chile R-123 (Meissner *et al.*, 2002) and the partial genomic sequences of isolate AH-1 (López *et al.*, 1997; Padula *et al.,* 2002). We report here the complete genomic sequence of the Chilean ANDV CHI-7913, whose isolation from a 10-year-old HPS patient has been recently described (Galeno, *et al.,* 2002). A preliminary account of this work has been reported (Tischler *et al.*, 2002).

Corresponding author: Pablo D.T. Valenzuela, Fundación Ciencia para la Vida, Av. Marathon 1943, Santiago, Chile. Telephone: (56-2) 239-8969. E-mail: pvalenzu@bionova.cl

MATERIALS AND METHODS

Virus, cells and media

The Andes virus used in this study was isolated from a 10-year-old patient who died of HPS (ANDV CHI-7913) and has been described previously (Galeno *et al.*, 2002). Vero E6 cells were originally provided by Dr. B. Hjelle, and were grown in Eagles minimal essential medium (MEM). The virus was grown via the conventional method (Lee, H.W., 1999) in Vero E6 cells grown to confluence in T25 flasks with MEM containing 10% fetal calf serum and antibiotics. The presence of viral antigens in infected cells was confirmed by immunofluorescence as described previously (Galeno *et al.*, 2002). All experiments involving infectious virus were performed in a bio-safety level 3 (BLS-3) laboratory.

Viral RNA preparation

Confluent mono-layers of Vero E6 cells were grown in MEM with 10% FCS and infected with ANDV isolate CHI-7913 and incubated at 37ºC in the presence of 5% CO_2. Cells were usually harvested two weeks after infection, lysed with a solution of guanidinium isothiocyanate (Cathala *et al.*, 1983) and the contents transferred to tubes. For RNA extraction, chloroform was added to the tubes, mixed and centrifuged at 12,000 xg for 15 min at 4º C. The RNA present in the aqueous phase was precipitated by the addition of 0.5 volumes of isopropanol. The pellet was washed twice with 70% ethanol, air dried, and dissolved in diethyl pyrocarbonate-treated water.

Genome segment amplification and sequencing

Approximately 1-4 µg of viral RNA was used in each reverse transcriptase reaction. The oligonucleotide primers for reverse transcription, as well as for the initial PCR reactions, were based on previously published well-conserved hantaviral sequences (Chizhikov *et al.*, 1995). Subsequently ANDV CHI-7913-specific primers were designed as sequence information became available. Primers were selected as required to cover the entire genome segments by overlaps. The amplified PCR products were purified with a purification kit (Wizard PCR Preps, Promega, Madison, WI, USA), ligated to the vector pGEM-T (Promega, Madison, USA), and used to transform *E. coli* DH5α cells. Clones with inserts of appropriate size were selected for sequencing using an ABI 310 instrument. To increase the accuracy of the sequence obtained, more than two clones from different PCR products were sequenced.

Nucleotide and amino acid sequence comparisons and analysis

The sequences of the fragments were assembled and edited into contigs using the Sequencher 4.1 and Vector NTI 6.1 programs. Nucleotide sequence analysis and comparisons were carried out using the alignment program BLASTN (NCBI). Protein alignments were realized with BLASTP (NCBI) and ClustalW, based on the Blosum62 substitution matrix. The hydrophobic domains were predicted by TmPred (EMBNet), while the N-glycosylation and O-glycosylation predictions were realized with NetNGlyc 1.0 and NetOGlyc 2.0 (CBS) respectively.

RESULTS AND DISCUSSION

Viral nucleotide sequences were determined by direct sequence analysis of RT-PCR products amplified from RNA preparations of Vero E6 cells infected with ANDV CHI-7913 (Galeno *et al.*, 2002). The nucleotide sequence has been deposited in the Gene Bank under the accession numbers AY228237, AY228238, and AY228239. The features of the nucleotide and amino acid sequence of the S, M and L viral RNA segments are described below.

S segment sequence properties and comparisons

The S RNA segment of the human Andes virus isolate CHI-7913 consists of 1,802 bases encoding a nucleocapsid protein of 428 amino acids. Additionally, the sequence analysis indicates the presence of 10 ORFs; the largest of them encoding a putative protein of 81 amino acids. The ORF in the + 1 frame encodes a putative protein of 63 amino acids, which could be a non-structural protein (NS) (Plyusnin, 2002), similar to the NS-S proteins of Bunyavirus, Phlebovirus and Tospovirus of the *Bunyaviridae* family (Elliott, 1990).

Comparisons of the nucleotide sequences of the S segment of the isolate ANDV CHI-7913 with ANDV Chile R123 and ANDV AH-1 reveals a sequence identity of 92% and 97% respectively, while only 83% similarity was found in comparison with the North American Sin Nombre virus (SNV), isolate NM H10. Alignment of the 504-base-long 3' non-coding region of CHI-7913 with the same region of ANDV AH-1 and ANDV Chile R123 reveals 97% and 89% identity respectively, while no significant similarity was found with the 3' non-coding region of SNV isolates. For Hantavirus it is known that their 3' non-coding region varies widely, both in nucleotide sequence and length (from 229 nucleotides in Prospect Hill virus to 728 nucleotides in SNV), except for the terminal nucleotides forming the panhandle structures. Assuming that the 3' non-coding region participates in steps of viral replication and packaging, there could be two possible explanations for the difference in their primary structures: (1) molecular mechanisms of replication differ from one host to another; or (2) the secondary rather than the primary structure of the 3' non-coding region is crucial for its proper activity (Plyusnin *et al.*, 1996).

The 428-amino-acid nucleoprotein encoded by the S segment belongs to the most conserved proteins of Hantavirus and has been proven to be highly immunogenic in humans (Jenison *et al.*, 1994, Bharadwaj *et al.*, 2000), thus making it the protein of choice for immunodiagnostic tests to detect virus infections. Despite its high similarity to ANDV (99% with isolates AH-1 and Chile R123) and SNV (93% with isolate NM H10) or other Hantavirus (e.g. 85% with Puumala isolate Hallnas B1 and 77% with Hantaan isolate 76–118), no conserved amino acid sequence motifs could be identified in comparisons done with nucleoproteins from Bunyavirus, Phlebovirus, Nairovirus and Tospovirus. The absence of a common motif supports the theory, that determined secondary, tertiary or even quaternary structures but not sequence specific interactions are responsible for RNA binding.

M segment sequence properties and comparisons

The largest ORF of the 3,641-base-long M segment RNA of isolate CHI-7913 encodes a glycoprotein precursor from nucleotide 34 to 3,450. Another 12 ORFs are present in the M segment, encoding putative proteins between 34 and 103 amino acids. The entire M segment shows 94% identity with the ANDV isolate Chile R123, 96% with AH-1, but only 77% with the SNV isolate NM H10. The 5' non-coding region proved to be with 100% identity highly conserved with the same region of ANDV isolates Chile R123 and AH-1, while the 3' non-coding region showed 91% and 92% identity respectively. In comparison with the 5' and 3' non-coding regions of SNV isolate NM H10, no significant similarities were found.

The 1,138-amino-acid-glycoprotein precursor encoded by the M segment RNA is the most variant protein of ANDV CHI-7913, presenting only an 87% similarity to the glycoprotein precursor of SNV isolate NM H10. When the amino acid sequence of the human isolate ANDV CHI-7913 was compared with the ANDV isolates Chile R123 and AH-1, nine non-conserved amino acid differences were found (see Table I).

Remarkable for Hantavirus glycoproteins is the high content of highly conserved cysteine residues (5.2%) and the virtually identical hydrophobic profiles (Antic, *et al.*, 1992), indicating a highly conserved tertiary protein structure of the hantaviral envelope proteins.

TABLE I

Non-conserved amino acid changes among different Andes virus isolates.
Conserved amino acids were defined by the Blosum62 substitution matrix.
The sign (–) indicates not available sequence information.

Genome	Encoded residue no.	ANDV CHI-7913	ANDV Chile R123	ANDV AH-1
S segment	148	A	T	T
M segment	23	T	I	I
	164	R	S	S
	281	R	I	R
	317	T	T	I
	328	L	L	F
	705	S	P	P
	732	T	I	I
	938	A	T	T
	1023	A	A	T
L segment	141	I	T	–
	271	L	F	–
	277	S	L	–
	319	P	L	–
	888	G	R	–
	1080	A	V	V
	1138	G	E	–
	1216	P	S	–
	1900	I	N	–
	2108	G	D	–
	2113	A	T	–
	2131	S	F	–

When Hantavirus glycoproteins are aligned with glycoproteins of Phlebovirus, Bunyavirus, Nairovirus and Tospovirus, only two of these cysteine residues (positions 768 and 773 in the CHI-7913 sequence) are maintained invariant. The cysteine residue at amino acid position 12 is only found in Hantavirus causing HPS, with the exception of Laguna Negra virus. The cysteine residues at positions 637 and 642 are associated only with South American Hantavirus (AND, Laguna Negra, Hu39694, Lechiguanas and Oran virus), while the cysteine at position 511 was found exclusively in ANDV (isolates CHI-7913, Chile R123 and AH-1). We have also found that these conserved cysteine residues are located in a highly conserved amino acid motif among members of the *Bunyaviridae* family (Table II). This motif (from amino acids 766 to 781) is located in the N-terminal part of the mature G2 protein and contains in addition to the conserved cysteines, partially conserved tryptophan and glycine residues.

It has been reported that vitronectin, which binds to ß3-integrins in an arginine-glycine-aspartic acid dependent manner, inhibits the entry of SNV and New York virus (NYV) into Vero E6 cells (Gavrilovskaya et al., 1998). We have found that ANDV, SNV and NYV share six conserved amino acid sequences with human and mouse vitronectin (Table III). Of these motifs, the motif 1 (amino acids 422 to 431) contains a highly conserved arginine and aspartic acid, while the motif 6 (amino acids 872 to 888) contains two highly conserved glycines.

The predicted glycoprotein precursor of the ANDV CHI-7913 contains four hydrophobic domains. The first domain extends from the N-terminal amino acids 1 to 19 and is most likely a hydrophobic signal peptide, which exists typically in secretory and type I trans-membrane proteins for their translocation into the lumen of the ER (Rapoport *et al.,* 1996). A signal peptidase complex usually removes the signal sequence (Perlman and Halvorson, 1983). This seems to be also valid for Hantavirus glycoproteins, as the

TABLE II

Conserved amino acid motif in the glycoprotein precursor of different
members of the *Bunyaviridae* family. Bold letters indicate identical,
underlined letters conserved amino acids.

Virus (strain)	Genus	Conserved Motif
La Crosse (Human/78)	Bunyavirus	1066 **WGC**EEFG**C**L**A**V**SD**G**CV**
BUNYWV (11573)	Bunyavirus	988 **WGC**EEFG**C**L**A**V**NT**G**CV**
SNV (NM H10)	Hantavirus	767 **WGC**NPPD**C**P**G**V**GT**G**CT**
ANDV (CHI-7913)	Hantavirus	766 **WGC**NPGD**C**P**G**V**GT**G**CT**
PUUV (Hallnas B1)	Hantavirus	769 **WGC**NPPD**C**P**G**V**GT**G**CT**
HTNV (76-118)	Hantavirus	763 **WGC**NPSD**C**P**G**V**GT**G**CT**
RVFV (ZH-548 M12)	Phlebovirus	821 **WGC**G---**C**FN**V**NP**S**C**L
PTPV (11587)	Phlebovirus	940 I**GC**G---**C**FN**I**NP**S**C**L
UUK (S23)	Phlebovirus	646 AL**C**Q---**C**FN**M**RP**S**C**F
TWSV (Brazilian BR-01)	Tospovirus	648 **WGC**EEAW**C**FA**I**NE**G**AT
CCHFV (Matin)	Nairovirus	1196 **W**R**C**NPTW**C**WG**V**GT**G**CT**

TABLE III

Conserved amino acid motifs shared by glycoprotein precursors (GP) of HPS
causing virus (ANDV, SNV, NYV) and the ß3-integrin ligand vitronectin in
comparison with nonpathogenic Hantavirus (PHV). Bold letters indicate identical
amino acids, underlined letters conserved amino acids.

Protein	Motif 1	Motif 2	Motif 3
GP PHV (11603)	426 Q**I**KFVCQ**RVD**	758 FLEKDYEF	942 NS**LEW**V**D**P
GP ANDV (CHI-7913)	422 K**I**N**F**ICQ**RVD**	755 **F**FEKDY**QY**	939 NK**LEW**I**D**P
GP SNV (NM H10)	423 R**I**N**FM**CQ**RVD**	756 **F**FEKDY**QY**	940 NR**LEW**I**D**P
GP NYV (NY-1)	424 R**I**N**FM**CQ**RVD**	756 **F**FEKDY**QY**	940 NR**LEW**I**D**P
Vitronectin human (SGHU1V)	446 R**V**N**L**RTR**RVD**	272 **Y**FFKGK**QY**	420 YR**M**D**W**L**V**P
Vitronectin mouse (SGMSV)	445 R**V**N**L**RTR**RVD**	270 **Y**FFKGK**QY**	420 YD**M**D**W**L**V**P

Protein	Motif 4	Motif 5	Motif 6
GP PHV (11603)	575 MC**PYC**MNP	588 **SAL**Q**AH**FK	875 **F**G**DPGD**IM**S**TP**Q**GM**Q**CP
GP ANDV (CHI-7913)	572 QC**PYC**M**T**I	585 **SAL**Q**AH**YS	872 **F**G**DPGD**IM**S**TP**S**GM**R**CP
GP SNV (NM H10)	573 QC**PYC**M**T**I	586 **SAL**Q**AH**FA	873 **F**G**DPGD**IM**S**TT**S**GM**R**CP
GP NYV (NY-1)	573 QC**PYC**M**T**M	586 **SA**L**Q**AH**FA	873 **F**G**DPGD**IM**S**TT**T**GM**K**CP
Vitronectin human (SGHU1V)	458 DP**PYC**PR**S**I	299 **SA**V**FEHFA**	200 **W**G**IEGPID**A**A**F**T**R**IN**CQ
Vitronectin mouse (SGMSV)	457 NP**PYC**PR**S**I	297 **SA**V**FEHFA**	199 **W**G**IEGPID**A**A**F**T**R**IN**CQ

exact N-terminus of the mature G1 protein of Hantaan virus as been determined to start at the position threonine 18 (Schmaljohn et al., 1987). Studies of B cell epitopes confirm this results also for ANDV, as sera of patients infected with ANDV revealed a strong reactivity with a peptide carrying amino acids 14–26, but not with a peptide carrying amino acids 1–13 (N. Tischler, unpublished data). The second and fourth hydrophobic domains (aa 490 to aa 509 and aa 1,109 to aa 1,129) are potential trans-membrane regions for each mature glycoprotein. The third hydrophobic region extends from amino acids 630 to 648 and is probably a second signal peptide for the cleavage of the glycoprotein precursor into the mature G1 and G2 proteins. The cleavage probably occurs at the highly conserved pentapeptide WAASA (Loeber *et al.,* 2001), also present in the ANDV CHI-7913 sequence.

Oligosaccharide chains are essential components for the correct folding of the glycoproteins and participate in many molecular recognition reactions of receptors and ligands. They can also determine whether viral envelope proteins are recognized by components of the host immune defense system. Hence, glycosylation differences could be related to pathogenic differences of virus strains.

Previous studies showed that only ANDV produce HPS synthomes in Syrian Hamsters while SNV does not (Hooper *et al.*, 2001). To investigate if those differences in pathogenesis could be realted with a different glycosylation pattern, the glycosylation prediction of amino acid sequences of 8 North American Hantavirus and 7 South American Hantavirus were compared. Five putative asparagine glycosylation sites with the characteristic sequence NXS/T (where X is any amino acid different to proline) are predicted for the ANDV CHI-7913 glycoprotein precursor. Of these, only four are conserved in HPS-associated Hantavirus (aa 138-141, aa 350-353, aa 402-405, aa 930-933). The fifth putative N-glycosylation site, which comprises the aa 524 527, was found in all of the South American Hantavirus studied, except in Laguna Negra virus. Within the

North American Hantavirus this site is present only in the Limestone Canyon and Prospect Hill virus (Table IV), of which no human infection has been determined (Schmaljohn *et al.*, 1997). Based on mucin-type O-glycosylation prediction, three main O-glycosylation clusters could be defined (Table IV). Among HPS associated Hantavirus, threonine 93 was found to be the only conserved putative O-glycosylation site, while threonine 92, threonine 583 and serine 307 are conserved in all of the South American Hantavirus. Interestingly, the glycosylation of serine 97 was predicted only for all South American Hantavirus, excluding Oran and Lechiguanas virus, which instead have an alanine residue. Additionally, the putative glycosylation cluster of serine 307, threonine 308 and serine 309 present in site 2 is found in all South American Hantavirus, with the exception of Laguna Negra virus, which presents an amino acid change of threonine 308 into serine. In contrast, the glycosylation of threonine 89 present in site 1 is predicted only for all North American Hantavirus, while threonine 90 is found to be glycosylated in all North American Hantavirus except Limestone Canyon and Bayou virus.

L segment sequence properties and comparisons

The 6,466-base-long L segment RNA of CHI-7913 contains a 6,453 base long ORF (nucleotide 9-6,464), which encodes a protein of 2,152 amino acids possessing at least the activity of a RNA dependent RNA polymerase. Additionally, another 21 ORFs, which encode putative proteins of 34 to 100 amino acids, are present. Nucleotide sequence comparison analysis with the L segment of another complete genome of ANDV (Chile R123) revealed 93% of identity, while the identity with the L segment of SNV virus isolate NM H10 amounted to 72%. Notably, in six regions of the SNV L segment no significant similarity was found (see Table V). However, when translating those regions into the corresponding amino acid sequence, similarities between 76% and 96% could be determined. This result emphasizes the

TABLE IV

O-glycosylation and N-glycosylation differences in the glycoprotein precursor of representative American Hantavirus. The numbers correspond to the first amino acid of the sequences. Bold characters indicate amino acids which are predicted to be glycosylated.

Virus (strain)	Similarity with CHI-7913 [%]	O-glycosylation Site 1	O-glycosylation Site 2	N-glycosylation difference	O-glycosylation Site 3
Sin Nombre (NM H10)	87	67 DLHVPATTYQKYNQVDWTKKSS**TTEST**NAGATTFEAKTK	301 PVTAKVP**S**T**T**ET**M**	519 SCSHYS**T**ESKFKVIL	578 MTI**T**ES**T**ESALQAH
Convict Creek (74)	83	67 DLHVPATTYQKYNQVDWTKKSS**TTEST**NAGA**T**TFEAKTK	301 PVTAKVP**S**T**T**ET**M**	519 SCSHYS**T**ESKFKVIL	578 MTI**T**ES**T**ESALQAH
New York (NY-2)	83	67 DLHVPSTSIQKYNQVEWAKKSS**TTEST**SAGATTFEAKTK	301 PITAKVP**S**T**T**E**AM**	519 SCSHYS**T**ESKFKAIL	578 MTM**T**ES**T**ESALQAH
Blue River (Oklahama)	83	67 DLHVPSTTIQKYTQVEWAKKSS**TTEST**NAAATTFEAKTK	301 PITAKVP**S**T**T**ET**M**	519 SCSHYSKESRFKILL	578 MTM**T**ES**T**ESALQAH
Limestone Canyon (68273)	82	42 DLHVPSTAIQKFTQVEWVKKTST**T**D**ST**NAAS**S**TFE**T**K**SK**	270 PINAKVPHTE**T**SDTL	488 SCSHYN**N**ESKFKLIL	547 MTL**T**ES**T**ESALQAH
Moro Canyon (RM-97)	81	67 DLHSPSNTVQKFTQVEWSKKAS**TT**D**ST**NAATTFE**T**KSK	301 LIAGKVPH**T**ET**T**ETF	519 SCSHYSSESKFKFIL	578 MTL**T**ES**T**ESALQAH
Bayou (37962)	83	67 DHHAPP**TT**AQ**T**F**TQ**L**S**W**SKTA**S**T**T**D**TT**NAGENTFE**T**KSA	303 PIKGKVPH**T**ET**T**ETV	521 SCSHYS**T**ESKFKIIL	580 MTL**T**E**AT**ESALQAH
Andes (CHI-7913)	100	67 DLH**T**TSMAQKSFTQVEWRKKSD**TT**D**TT**NAA**S**TTFEAQTK	300 PITAKVP**S**T**S**STDTL	518 SCSHY**T**NESKFKFIL	577 MTI**T**E**AT**ESALQAH
Andes (AH-1)	98	67 DLH**T**TSMAQKSFTQVEWRKKSD**TT**D**TT**NAA**S**TTFEAQTK	300 PITAKVP**S**T**S**STDTL	518 SCSHY**T**NESKFKFIL	577 MTI**T**E**AT**ESALQAH
Andes (Chile R123)	98	67 DLH**T**TSMAQKSFTQVEWRKKSD**TT**D**TT**NAA**S**TTFEAQTK	300 PITAKVP**S**T**S**TDTL	518 SCSHY**T**NESKFKFIL	577 MTI**T**E**AT**ESALQAH
Lechiguanas (OF22819)	92	67 DLH**T**TSM**T**QKSYTQVEWRKKSD**TT**D**TT**NAAATTFESQAK	300 PLTAKVP**S**T**S**SSDTL	518 SCSHYNESKFKFIL	577 MTM**T**E**AT**ESALQAH
Oran (OI22996)	91	67 DLY**T**TPM**T**QKSYTQVEWRKKSE**TT**D**TT**NAAATTFESQTK	300 PITAKVP**S**T**S**SSDTL	518 SCSHYNESKFKLIL	577 MTM**T**E**AT**ESALQAH
Hu39694 (hu39694)	91	67 DLH**T**TSM**T**QKSYTQVEWRKKSD**TT**D**TT**NAA**S**TTFEAQTK	300 PITAKVP**S**T**S**STDTL	518 SCSHY**T**NESKFKFIL	577 MTM**T**E**AT**ESALQAH
Laguna Negra (510B)	89	67 DLH**T**TS**S**TQKSFTQLEWRKKAE**TT**D**TT**Q**A**AS**T**TFEAQTK	300 PITAKVP**S**SSS**T**ETL	518 SCSHYSSESKFKMIL	577 MTV**T**E**AT**QSALQAH
Prospect Hill (11603)	79	68 DVH**T**SSATQQAVTKWTWEKKAD**T**AE**T**AKAAS**T**TFQSKST	303 KLSGKAP**S**TE**S**SDTV	521 MCSKYS**N**DSKFRLLI	576 MNP**T**ES**T**ESALQAH

TABLE V

Different codon usage of AND and SN virus L segments. Alignments between L segments
and encoded proteins of ANDV isolates CHI-7913, Chile R123 and SN virus NM H10. The
sign (-) indicates no significant similarity (E value > 10).

Nucleotide sequence of the L segment of ANDV CHI-7913	Sequence comparisons			
	ANDV Chile R123	SNV NM H10		
	Nucleotide identity (%)	Nucleotide identity (%)	Encoded amino acids	
			Identity (%)	Similarity (%)
417 – 497	91	-	69	76
595 – 917	91	-	70	85
1186 – 1336	98	-	62	84
3878 – 3962	95	-	89	92
4574 – 4862	94	-	87	96
6314 – 6465	94	-	66	84

difference of codon usage in these regions by the ANDV and SNV. However, the conservation of the predicted primary structure of the protein, despite extreme mutations in the nucleotide sequence, could be a result of evolutionally adaptation of the virus to the host codon usage.

The protein of 2,152 amino acids encoded by the L segment of CHI-7913 shows an overall identity of 98% with ANDV R123, while the similarity with SNV NM H10 amounts to 94%. When the derived amino acid sequence of the human isolate CHI-7913 was compared with isolates Chile R123 (complete sequence) and AH-1 (sequence from aa 933 to 1134), twelve non-conserved amino acid changes were found (see Table I).

The deduced amino acid sequence of ANDV CHI-7913 RNA polymerase contains the six motifs previously identified as common to all RNA-dependent RNA polymerases (Poch *et al.*, 1989; Müller *et al.*, 1994), which have been suggested to be related to polymerase function (Müller *et al.*, 1994). All of these motifs are located in the most highly conserved middle region of the protein (from amino acids 860 to 1180). A sequence alignment with other members of the *Bunyaviridae* family is shown in Table VI. Motif preA characterized by the presence of three strictly conserved basic residues is present between lysine 885 and

glutamic acid 910 of ANDV CHI-7913. The first two conserved basic residues of this motif, lysine 885 and arginine 893, are probably part of the "finger sub-domain" of a proposed "right hand" structural model of polymerases (Ollis *et al.*, 1985; Kohlstaedt et al., 1992) and discussed in more detail by Hansen *et al.*, (1997) and O'Really and Kao, (1998). These two residues may participate in positioning and binding. It has been suggested that the third positively charged residue, arginine 903 in ANDV CHI-7913 is located in the vicinity of the active site, close to the binding site of the template strand (Müller et al., 1994).

Motif A, located between residues lysine 964 and asparagine 982 in ANDV CHI-7913 and motif C, located between phenylalanine 1103 and asparagine 1116 in ANDV CHI-7913 (Table VI), are characterized by the presence of strictly conserved aspartic acid residues which have been suggested to be part of the active site (Delarue *et al.*, 1990) and to participate in the binding of metal ions involved in catalysis by the Klenow fragment (Beeze and Steitz, 1991). Motif B, located between residues alanine 1053 and alanine 1074 of ANDV CHI-7913 (Table VI), is characterized by the presence of a strictly conserved glycine residue and may also participate in template binding (Müller *et al.,* 1994). Motifs D and E, located between glycine 1154 and phenylalanine

TABLE VI

Conserved amino acid motifs among polymerases of different genus
members of the *Bunyaviridae* family. Bold letters indicate identical,
underlined letters conserved amino acids.

Virus (strain)	Genus	Pre-Motif A	Motif A
La Crosse (Human/78)	Bunyavirus	950 **KGQ**KTSKD**R**EIF**V**GEYEA**K**MCMYA**VE**	1051 KGLKMEINA**D**MS**KW**SAQDV
BUNYWV (11573)	Bunyavirus	881 **KGQ**KTAKD**R**EIF**V**GEFEA**K**MCMYV**VE**	958 KALKLEINA**D**MS**KW**SAQDV
SNV (NM H10)	Hantavirus	885 **KYQ**RTEAD**R**GFF**I**TTLPT**R**VRLEII**E**	964 KRRLMYVSA**D**AT**KW**SPGDN
ANDV (CHI-7913)	Hantavirus	885 **KYQ**GTEAD**R**GFF**I**TTLPT**R**VRLEII**E**	964 KRKLMYVSA**D**AT**KW**SPGDN
PUUV (Hallnas B1)	Hantavirus	814 **KYQ**RTEAD**R**GFF**I**TTLPT**R**VRLEII**E**	893 KRKLMYVSA**D**AT**KW**SPGDN
HTNV (76–118)	Hantavirus	814 **KYQ**RTEAD**R**GFF**I**TTLPT**R**CRLEII**E**	893 KRKLMYVSA**D**AT**KW**SPGDN
RVFV (ZH-548 M12)	Phlebovirus	848 **K**KQQHGGL**R**EIY**V**MGAEE**R**IVQSV**VE**	910 PVWTCATSD**D**AR**KW**NQGHF
TOSV (ISS.PHL.3)	Phlebovirus	849 **K**KQQHGGL**R**EIY**V**MGADE**R**IVQSV**IE**	913 SVWTCATSD**D**AR**KW**NQGHY
UUKV (S23)	Phlebovirus	852 **K**PQ-HGGL**R**EIY**V**LGFEE**R**VVQLV**IE**	916 HHETVATSD**D**AA**KW**NQCHH
TWSV (Brazilian BR-01)	Tospovirus	1282 **KMQ**RTKTD**R**EIY**L**MSMKV**K**MMLYF**IE**	1350 KSRLAFLSA**D**QS**KW**SASGL

Virus (strain)	Motif B	Motif C	Motif D and E
La Crosse (Human/78)	1139 VLIKRNWL**QG**NFNYT**SS**YVHCS	1178 ILVNSLVHS**DD**NQT	1220 FGCQANM**KK**T**Y**V**T**NCIK**E**FV**S**LF
BUNYWV (11573)	1046 VQIKRNWL**QG**NFNYI**SS**YVHSC	1085 CLINSMVHS**DD**NQT	1127 FGCQANM**KK**T**Y**I**T**HTCK**E**FV**S**LF
SNV (NM H10)	1053 ATVKGNWL**QG**NLNKC**SS**LFGAA	1103 FIYGYLEPE**DD**GTD	1154 GSIKVSP**KK**T**T**V**S**PTNA**E**FL**S**TF
ANDV (CHI-7913)	1053 ANVKGNWL**QG**NLNKC**SS**LFGAA	1103 FIYGYLEPE**DD**GTD	1154 GSIKVSP**KK**T**T**V**S**PTNA**E**FL**S**TF
PUUV (Hallnas B1)	982 ASIKGNWL**QG**NLNKC**SS**LFGAA	1032 FIYGYLEPV**DD**GTE	1083 GSIKISP**KK**T**T**V**S**PTNA**E**FL**S**TF
HTNV (76–118)	982 GEVKGNWL**QG**NLNKC**SS**LFGVA	1032 FIYGYLEPV**DD**GTD	1083 GSIKISP**KK**T**T**V**S**PTNA**E**FL**S**TF
RVFV (ZH-548 M12)	1008 LETTTGMM**QG**ILHYT**SS**LLHTI	1054 LVCDMMQGS**DD**SSM	1100 LAIYPSE**K**S**T**AN**T**DFVM**E**YN**S**EF
TOSV (ISS.PHL.3)	1009 LKTSTGMM**QG**ILHFT**SS**LLHSL	1055 VVCDMMQGS**DD**SSM	1101 IGIYPSE**K**S**T**PN**T**DFVM**E**YN**S**EF
UUKV (S23)	1013 VQTETGMM**QG**ILHYT**SS**LLHTL	1059 VLVDVLQSS**DD**SGM	1105 LGIYSS**VK**S**T**NN**T**LHLL**E**FN**S**EF
TWSV (Brazilian BR-01)	1444 YPVSMNWL**QG**NLNYL**SS**VYHSC	1482 FQTRWIVHS**DD**NAT	1530 FCITLNP**KK**S**Y**A**S**SSEV**E**FI**S**ER

1176 in ANDV CHI-7913 (Table VI), are characterized by the presence of strictly conserved glycine, glutamic acid and serine residues. Motifs preA, A, B, C, D and E have been associated to the "right hand" structural model of RNA polymerases of which motifs A, B, C and D are present in all polymerases, while motif preA is only found in RNA dependent RNA polymerases and motif E only in polymerases of segmented, negative-stranded RNA virus (Hansen *et al.,* 1997; O'Really and Kao, 1998).

The complete sequence of the genome and the comparative analysis of the predicted translated proteins of the human isolate ANDV CHI-7913 reported here represents additional information to those reported in the past (López *et al.,* 1997, Padula et al. 2002, Meissner *et al,* 2002). This will allow further sequence comparisons useful in future studies of ANDV pathogenicity and the development of immunodiagnostics, vaccines and potential immunotherapeutic agents.

special thanks to Danilo González of the University of Santiago de Chile for his valuable support in bioinformatics.

REFERENCES

ANTIC D, WRIGHT KE, KANG CY (1992) Maturation of Hantavirus proteins G1 and G2. Virol 189: 324-328

BEEZE L, STEITZ TA (1991) Structural basis for the 3'-5' exonuclease activity of *E. coli* DNA polymerase I: a two metal ion mechanism. EMBO J 10:25-33

BHARADWAJ M., NOFCHISSEY R., GOADE D, KOSTER F, HJELLE B (2000) Humoral immune responses in the Hantavirus cardiopulmonary syndrome. J Inf Dis 182:43-48

CATHALA G, SAVOURET F, MÉNDEZ B, WEST B, KARIN M, MARTIAL J, BAXTER J (1983) A method for isolation of intact, translationally active ribonucleic acid. DNA 2(4):329-35

CHIZHIKOV VE, SPIROPOULOU CF, MORZUNOV SP, MONROE MC, PETERS CJ, NICHOL ST (1995) Complete genetic characterization and analysis of isolation of *Sin Nombre* Virus. J Virol 69:8132-8136

DELARUE M, POCH O, TORDO N, MORAS D, ARGOS P (1990) An attempt to unify the structure of polymerases. Protein Eng 3: 461-467

ELLIOTT RM (1990) Molecular biology of the *Bunyaviridae*. J Gen Virol 71: 501-522

GAVRILOVSKAYA IN, SHEPLEY M, SHAW R, GINSBERG M, MACKOW E (1998) ß3 integrins mediate the cellular entry of Hantaviruses that cause respiratory failure. Proc Natl Acad Sci USA 95: 7074-7079

GALENO H., MORA J, VILLAGRA E, FERNÁNDEZ J, HERNÁNDEZ J, MERTZ GJ, RAMÍREZ E (2002) First human isolate of Hantavirus (Andes virus) in the Americas. Emerg Infect Dis 8: 657-661

HANSEN JL, LONG AM, SCHULTZ SC (1997) Structure of the RNA dependent RNA polymerase of poliovirus. Current Biol Struct 5: 1109-1122

HOOPER JW, LARSEN T, CUSTER DM, SCHMALJOHN CS (2001) A lethal disease model for hantavirus pulmonary syndrome, Virology 10; 289(1): 6-14

JENISON S, YAMADA T, MORRIS C, ANDERSON B, TORREZ-MARTÍNEZ N, KELLER N, HJELLE B (1994) Characterization of human antibody responses to Four Corners Hantavirus infections among patients with Hantavirus pulmonary syndrome. J Virol 68: 3000-3006

KOHLSTAEDT LA, WANG J, FRIEDMAN JM, RICE PA, STEITZ TA (1992) Crystal structure at 35 A° resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Sci 256 : 1783-1790

LEE HW (1999) Virus isolation. In: LEE HW, CALISHER C, SCHMALJOHN C (Eds). Manual of hemorrhagic fever with renal syndrome and Hantavirus pulmonary syndrome. Seoul: WHO Collaborating Centre for Virus Reference and Research (Hantavirus). pp:74-79

LOEBER C, ANHEILER B, LINDOW S, KLENK HD, FELDMANN H (2001) The Hantaan virus glycoprotein precursor is cleaved at the conserved pentapeptide WAASA: Virol 289: 224-229

LÓPEZ N, PADULA P, ROSSI C, LÁZARO ME, FRANZE-FERNÁNDEZ MT (1996) Genetic identification of a new Hantavirus causing severe pulmonary syndrome in Argentina. Virol 220: 223–226

LÓPEZ N, PADULA P, ROSSI C, MIGUEL S, EDELSTEIN A, RAMÍREZ E, FRANZE-FERNÁNDEZ MT (1997) Genetic characterization and phylogeny of Andes virus and variants from Argentina and Chile. Virus Res 50: 77–84

MEISSNER JD, ROWE JE, BORUCKI MK, STEPHEN CSJ (2002) Complete nucleotide sequence of a Chilean Hantavirus. Virus Res 89: 131-143

MÜLLER R, POCH O, DELARUE M, BISHOP DHL, BOULOY M (1994) Rift Valley fever virus L segment: correction of the sequence and possible role of newly identified regions conserved in RNA-dependent polymerases. J Gen Virol 75: 1345-1352

OLLIS DL, BRICK P, HAMLIN R, XUONG NG, STEITZ TA (1985) Structure of the large fragment of *E. coli* DNA polymerase I complexed with dTMP. Nature 343:762-766

OREILLY E.K, KAO CC (1998) Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. Virol 252: 287-303

PADULA PJ, SÁNCHEZ AJ, EDELSTEIN A, NICHOL ST (2002) Complete nucleotide sequence of the M RNA of Andes virus and analysis of the variability of the termini of the virus S, M and L RNA segments. J Gen Virol 83: 2117-2122

PERLMAN D, HALVORSON HO (1983) A putative signal peptidase recognition site and sequence in eucaryotic and procaryotic signal peptides. J Mol Biol 167: 391-409

PLYUSNIN A, VAPALAHTI O, VAHERI A (1996) Hantavirus: genome structure, expression and evolution. J Gen Virol 77: 2677–2687

PLYUSNIN, A (2002) Genetics of hantaviruses: implications to taxonomy, Arch. Virol. 147: 665-682

POCH O, SAUVAGET I, DELARUE M, TORDO N (1989) Identification of four conserved motifs among the RNA-dependent polymerase elements EMBO J 8: 3867-3874

RAPOPORT TA, JUNGNICKEL B, KUTAY U (1996) Protein transport across the eucaryotic endoplasmatic reticulum and bacterial inner membranes. Ann Rev Biochem 65: 271–303

SCHMALJOHN C, HJELLE B (1997) Hantaviruses: A global disease problem. Emerg Infect Dis 3: 95-104

SCHMALJOHN CS, SCHMALJOHN AL, DALRYMPLE JM (1987) Hantaan virus M RNA: Coding strategy, nucleotide sequence, and gene order. Virol 157, 31–39

TISCHLER N, LISBONA F, MÜLLER I, MARTÍNEZ R, FERNÁNDEZ J, ROSEMBLATT M, VALENZUELA P (2002) Secuencia y expresión de genes de virus Hanta Andes. Biol Res 35: R-104