Volume 11 Number 17 1983 Nucleic Acids Research

Amino acid sequence of human respiratory syncytial virus nucleocapsid protein

Narayanasamy Elango and Sundararajan Venkatesan

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20205, USA

Received 15 June 1983; Accepted 3 August 1983

ABSTRACT

Amino acid sequence of the human respiratory syncytial (RS) virus nucleocapsid (NC) protein, deduced from the DNA sequence of a recombinant plasmid, is presented. The cDNA plasmid (pRSB11) has 1412 bp of RS viral NC sequence and lacks six nucleotides of the 5'end of mRNA. There is a single long open reading frame encoding 467 amino acids. This 51540 dal protein is rich in basic amino acids and has no homologies with other known viral capsid proteins.

INTRODUCTION

Human respiratory syncytial virus (RS virus), an unsegmented negative stranded RNA virus, is a leading etiological agent of severe lower respiratory tract illness among pediatric population (1). Although previously classified as a Paramyxovirus, it has been placed in a separate genus, Pneumovirus, on the basis of its morphology and lack of haemagglutinin and neuraminidase activities. The virus has an encapsidated 5000 kdal genomic RNA of negative polarity (2). The proteins associated with the virus include two surface glycoproteins (84 kdal and 68 kdal), a 46 kdal capsid protein (NC), a 36 kdal phosphoprotein (P), a 28 kdal matrix protein (M) and a 160 kdal protein presumed to be the viral polymerase. Two or three nonstructural proteins have been observed in infected cells (3,4).

Unsegmented negative stranded RNA viruses replicate in the cytoplasm of infected cells and possess a transcription system that generates, in a regulated manner, distinct virus specific mRNAs with typical eukaryotic features including poly(A) tail at the 3'end and a methylated cap structure at the 5'end (5). *In vitro* transcriptional studies with detergent solubilized virus (6) or infected cell extracts have established the viral nucleocapsid as the irreducible minimum for both transcriptional and replicative processes (7 and references therein). In the best studied

prototype of this class of of viruses, namely, vesicular stomatitis virus (VSV), the transcriptionally active nucleocapsid consists of a genomic RNA of about 12 kb in size, a 48 kdal capsid protein (N), a 180 kdal polymerase (L) and an additional 25 kdal protein (NS) (8). Paramyxovirus nucleocapsids are similar except having a larger genomic RNA (14-15 kb) and a phosphoprotein (P) which is the counterpart of the NS protein of VSV (9).

Substantial evidence exists relating to the transcriptional mechanism of this class of viruses. Typically, a small untranslated leader RNA, complementary to the 3'end of the genome, is synthesized initially followed by sequential synthesis of poly(A) containing mRNAs encoding different viral proteins (10). In all cases investigated so far, the most proximal mRNA synthesized is the one encoding the viral capsid protein (6,11,12). In the case of VSV, the switch from the transcriptional to replicative mode is dependent on the availability of N protein (13,14,15). The replicative process yields encapsidated positive stranded genome and not naked RNA. This has led to a model wherein the viral capsid protein interacts with a specific sequence within the leader RNA thereby relieving the polymerase from a possible attenuation signal and allowing the synthesis of full length copy of the genomic RNA (16). Indeed, it has been shown that both the positive and negative leader RNAs are present in the infected cells as nucleocapsids (17). Within the leader RNA, a specific sequence has recently been identified to be the encapsidation signal (14,18). This sequence is not present in viral mRNAs and hence they remain virtually unencapsidated. However, the regions on the protein responsible for the encapsidation process have not yet been mapped.

RS virus would be expected to share many of the properties described above. However, RS viral nucleocapsids are morphologically distinct (19) and have not been shown to be transcriptionally active. This led us to enquire whether there is a fundamental difference in the nucleocapsid assembly with this virus. Studies relating to the functional roles of the capsid protein, especially the manner in which it interacts with the genomic RNA and transcriptional enzymes, would be greatly facilitated by a knowledge of the primary structure of the protein. Here we report the polypeptide sequence of the RS viral capsid protein deduced from the DNA sequence of a recombinant plasmid previously identified to be containing this gene (20). The general features of this sequence are discussed.

MATERIALS AND METHODS

ENZYMES AND CHEMICALS

All restriction enzymes were purchased from either BRL or New England Biolabs. Boehringer Mannheim was the source of Klenow fragment of E. coli DNA polymerase I and calf intestinal alkaline phosphatase. Avian myeloblastosis virus reverse transcriptase was purchased from Life Sciences Corporation. Ribo and deoxyribonucleoside triphosphates and T4 polynucleotide kinase were obtained from P.L. Biochemicals. Radioactive nucleotides were from Amersham.

CELLS AND VIRUS

RS virus (strain A_2) was propagated in Hep-2 cell monolayers in Eagle's minimum essential medium supplemented with 2% heat inactivated fetal calf serum (20).

mRNA ISOLATION

Poly(A) containing RNA from RS virus infected cells, mock infected cells treated with actinomycin D (2 ug/ml) and from uninfected cells was isolated as described before (20).

cDNA SEQUENCING

A recombinant plasmid (pRSB11) containing the largest cDNA insert encoding the RS virus NC gene (20) was used in this study. Plasmid DNA was prepared using a modified alkali/SDS procedure (21) and a 1588 bp cDNA insert, lacking HpaII sites, was obtained by digesting the plasmid DNA with HpaII which cleaves the vector, pBR322, about 50 bp on either side of the PstI cloning site. Restriction enzyme sites were mapped by partial digestions of asymmetrically end labeled fragments (22) and by double digestions (23).

DNA sequencing was accomplished by the chemical method of Maxam and Gilbert (24). The sequencing reaction products were resolved under denaturing conditions on 5, 8 or 20% acrylamide gels in 8M urea. T4 polynucleotide kinase was used to label 5'ends and Klenow fragment of E. coli DNA polymerase was used to label 3'ends. Asymmetrically labeled DNA fragments, generated by secondary restriction enzyme digestions, were purified by DEAE paper electroelution following agarose gel electrophoresis (25). DNA sequence was analysed by the computer program of Queen and Korn (26). For hybridization purposes, uniformly labeled DNA was prepared by the nick translation procedure (27).

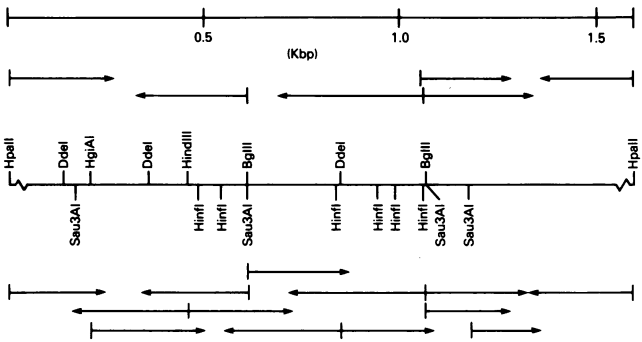

Fig. 1: Restriction map and sequencing strategy. The cDNA insert is represented by a straight line. The wavy lines on either side represent the pBR322 sequence as well as the G:C and A:T tails generated during the cloning procedure. The two HpaII sites occur approximately 50 bp on either side of the PstI site of pBR322 that was lost during cDNA construction. Relevant restriction enzyme sites within the insert are indicated. The arrows below and above illustrate the various 5' or 3' end labeled DNA fragments used for DNA sequencing. The bar represents the labeled site and the arrow head delineates the extent of sequence determination.

RESULTS

SEQUENCE OF RS VIRUS NUCLEOCAPSID PROTEIN

Previously we described the construction of cDNA clones for four RS viral genes (20). The largest recombinant plasmid (pRSB11), encoding the viral capsid protein, was devoid of the PstI sites. However, an insert of 1588 bp was obtained by digestion with HpaII which cleaves the vector, pBR322, about 50 bp on either side of the PstI site but not within the cDNA insert. This cDNA insert was then used to map relevant restriction sites. Fig. 1 illustrates the restriction map and the sequencing strategy.

The DNA sequence of the capsid gene and the translated sequence of the protein are presented in Fig. 2. There is a single long open reading frame starting at the twelfth nucleotide following the G:C tails upto position 1412. The other two reading frames are extensively blocked throughout. The translated amino acid sequence yields a polypeptide of 51540 daltons containing 467 amino acids. The protein is relatively rich in basic amino acids (Table I). It is poor in cysteine (3 residues) and relatively rich in proline (21 residues). The termination codon is six bases upstream of the poly(A) tail. In this sequence, there is no counterpart of the canonical eukaryotic polyadenylation signal, AAUAAA (28). The sequence AAGAUGG, flanking and including the initiator AUG sequence, is reminiscent of the consensus PXXAUGG sequence around functional eukaryotic

```
                                                    7                                                  17
CAAATACAAAG ATG GCT CTT AGC AAA GTC AAG TTG AAT GAT ACA CTC AAC AAA GAT CAA CTT
            MET ALA LEU SER LYS VAL LYS LEU ASN ASP THR LEU ASN LYS ASP GLN LEU
                                                   27                                                  37
CTG TCA TCC AGC AAA TAC ACC ATC CAA CGG AGC ACA GGA GAT AGT ATT GAT ACT CCT AAT
LEU SER SER SER LYS TYR THR ILE GLN ARG SER THR GLY ASP SER ILE ASP THR PRO ASN
                                                   47                                                  57
TAT GAT GTG CAG AAA CAC ATC AAT AAG TTA TGT GGC ATG TTA TTA ATC ACA GAA GAT GCT
TYR ASP VAL GLN LYS HIS ILE ASN LYS LEU CYS GLY MET LEU LEU ILE THR GLU ASP ALA
                                                   67                                                  77
AAT CAT AAA TTC ACT GGG TTA ATA GGT ATG TTA TAT GCG ATG TCT AGG TTA GGA AGA GAA
ASN HIS LYS PHE THR GLY LEU ILE GLY MET LEU TYR ALA MET SER ARG LEU GLY ARG GLU
                                                   87                                                  97
GAC ACC ATA AAA ATA CTC AGA GAT GCG GGA TAT CAT GTA AAA GCA AAT GGA GTA GAT GTA
ASP THR ILE LYS ILE LEU ARG ASP ALA GLY TYR HIS VAL LYS ALA ASN GLY VAL ASP VAL
                                                  107                                                 117
ACA ACA CAT CGT CAA GAC ATT AAT GGA AAA GAA ATG AAA TTT GAA GTG TTA ACA TTG GCA
THR THR HIS ARG GLN ASP ILE ASN GLY LYS GLU MET LYS PHE GLU VAL LEU THR LEU ALA
                                                  127                                                 137
AGC TTA ACA ACT GAA ATT CAA ATC AAC ATT GAG ATA GAA TCT AGA AAA TCC TAC AAA AAA
SER LEU THR THR GLU ILE GLN ILE ASN ILE GLU ILE GLU SER ARG LYS SER TYR LYS LYS
                                                  147                                                 157
ATG CTA AAA GAA ATG GGA GAG GTA GCT CCA GAA TAC AGG CAT GAC TCT CCT GAT TGT GGG
MET LEU LYS GLU MET GLY GLU VAL ALA PRO GLU TYR ARG HIS ASP SER PRO ASP CYS GLY
                                                  167                                                 177
ATG ATA ATA TTA TGT ATA GCA GCA TTA GTA ATA ACT AAA TTA GCA GCA GGG GAC AGA TCT
MET ILE ILE LEU CYS ILE ALA ALA LEU VAL ILE THR LYS LEU ALA ALA GLY ASP ARG SER
                                                  187                                                 197
GGT CTT ACA GCC GTG ATT AGG AGA GCT AAT AAT GTC CTA AAA AAT GAA ATG AAA CGT TAC
GLY LEU THR ALA VAL ILE ARG ARG ALA ASN ASN VAL LEU LYS ASN GLU MET LYS ARG TYR
                                                  207                                                 217
AAA GGC TTA CTA CCC AAG GAC ATA GCC AAC AGC TTC TAT GAA GTG TTT GAA AAA CAT CCC
LYS GLY LEU LEU PRO LYS ASP ILE ALA ASN SER PHE TYR GLU VAL PHE GLU LYS HIS PRO
                                                  227                                                 237
CAC TTT ATA GAT GTT TTT GTT CAT TTT GGT ATA GCA CAA TCT TCT ACC AGA GGT GGC AGT
HIS PHE ILE ASP VAL PHE VAL HIS PHE GLY ILE ALA GLN SER SER THR ARG GLY GLY SER
                                                  247                                                 257
AGA GTT GAA GGG ATT TTT GCA GGA TTG TTT ATG AAT GCC TAT GGT GCA GGG CAA GTG ATG
ARG VAL GLU GLY ILE PHE ALA GLY LEU PHE MET ASN ALA TYR GLY ALA GLY GLN VAL MET
                                                  267                                                 277
TTA CGG TGG GGA GTC TTA GCA AAA TCA GTT AAA AAT ATT ATG TTA GGA CAT GCT AGT GTG
LEU ARG TRP GLY VAL LEU ALA LYS SER VAL LYS ASN ILE MET LEU GLY HIS ALA SER VAL
                                                  287                                                 297
CAA GCA GAA ATG GAA CAA GTT GTT GAG GTT TAT GAA TAT GCC CAA AAA TTG GGT GGT GAA
GLN ALA GLU MET GLU GLN VAL VAL GLU VAL TYR GLU TYR ALA GLN LYS LEU GLY GLY GLU
                                                  307                                                 317
GCA GGA TTC TAC CAT ATA TTG AAC AAC CCA AAA GCA TCA TTA TTA TCT TTG ACT CAA TTT
ALA GLY PHE TYR HIS ILE LEU ASN ASN PRO LYS ALA SER LEU LEU SER LEU THR GLN PHE
                                                  327                                                 337
CCT CAC TTC TCC AGT GTA GTA TTA GGC AAT GCT GCT GGC CTA GGC ATA ATG GGA GAG TAC
PRO HIS PHE SER SER VAL VAL LEU GLY ASN ALA ALA GLY LEU GLY ILE MET GLY GLU TYR
                                                  347                                                 357
AGA GGT ACA CCG AGG AAT CAA GAT CTA TAT GAT GCA GCA AAG GCA TAT GCT GAA CAA CTC
ARG GLY THR PRO ARG ASN GLN ASP LEU TYR ASP ALA ALA LYS ALA TYR ALA GLU GLN LEU
                                                  367                                                 377
AAA GAA AAT GGT GTG ATT AAC TAC AGT GTA CTA GAC TTG ACA GCA GAA GAA CTA ACC CTC
LYS GLU ASN GLY VAL ILE ASN TYR SER VAL LEU ASP LEU THR ALA GLU GLU LEU THR LEU
                                                  387                                                 397
AAG ACA ACC AAA AAA GAT CCC AAA CCT CAA ACC ACT AAA TCA AAG GAA GTA CCC ACC ACC
LYS THR THR LYS LYS ASP PRO LYS PRO GLN THR THR LYS SER LYS GLU VAL PRO THR THR
                                                  407                                                 417
AAG CCC ACA GAA GAG CCA ACC ATC AAC ACC ACC AAA ACA AAC ATC ATA ACT ACA CTA CTC
LYS PRO THR GLU GLU PRO THR ILE ASN THR THR LYS THR ASN ILE ILE THR THR LEU LEU
                                                  427                                                 437
ACC TCC AAC ACC ACA GGA AAT CCA GAA CTC ACA AGT CAA ATG GAA ACC TTC CAC TCA ACT
THR SER ASN THR THR GLY ASN PRO GLU LEU THR SER GLN MET GLU THR PHE HIS SER THR
                                                  447                                                 457
TCC TCC GAA GGC AAT CCA AGC CCT TCT CAA GTC TCT ACA ACA TCC GAG TAC CCA TCA CAA
SER SER GLU GLY ASN PRO SER PRO SER GLN VAL SER THR THR SER GLU TYR PRO SER GLN
                                                  467
CCT TCA TCT CCA CCC AAC ACA CCA CGC CAG TAG TTACTTAAAAAAAAAAAAAAAAAAAAA
PRO SER SER PRO PRO ASN THR PRO ARG GLN END

MOLECULAR WEIGHT =  51540
```

Fig. 2: Sequence of RS virus capsid gene. The DNA sequence represents the messenger strand. The longest open reading frame is numbered starting with the N-terminal methionine.

translational initiation sites (29).

SIZE DETERMINATION OF CAPSID PROTEIN mRNA

Although the cDNA synthesis procedure used to construct the cDNA library was previouly shown to yield clones where the 5'end sequences of the mRNAs are preserved (30), we decided to determine whether this was true with pRSB11.

Table I: Codon usage and the amino acid composition of the nucleocapsid protein.

TTT	8	(1.7)	TCT	10	(2.1)	TAT	9	(1.9)	TGT	3	(0.6)
TTC	5	(1.1)	TCC	7	(1.5)	TAC	8	(1.7)	TGC	0	(0.0)
TTA	18	(3.8)	TCA	7	(1.5)	TAA	0	(0.0)	TGA	0	(0.0)
TTG	7	(1.5)	TCG	0	(0.0)	TAG	1	(0.2)	TGG	1	(0.2)
CTT	3	(0.6)	CCT	6	(1.3)	CAT	8	(1.7)	CGT	2	(0.4)
CTC	6	(1.3)	CCC	6	(1.3)	CAC	4	(0.9)	CGC	1	(0.2)
CTA	8	(1.7)	CCA	8	(1.7)	CAA	16	(3.4)	CGA	0	(0.0)
CTG	1	(0.2)	CCG	1	(0.2)	CAG	2	(0.4)	CGG	2	(0.4)
ATT	8	(1.7)	ACT	8	(1.7)	AAT	16	(3.4)	AGT	6	(1.3)
ATC	6	(1.3)	ACC	14	(3.0)	AAC	10	(2.1)	AGC	6	(1.3)
ATA	14	(3.0)	ACA	19	(4.1)	AAA	28	(6.0)	AGA	8	(1.7)
ATG	15	(3.2)	ACG	0	(0.0)	AAG	7	(1.5)	AGG	4	(0.9)
GTT	7	(1.5)	GCT	8	(1.7)	GAT	13	(2.8)	GGT	9	(1.9)
GTC	4	(0.9)	GCC	4	(0.9)	GAC	6	(1.3)	GGC	7	(1.5)
GTA	9	(1.9)	GCA	17	(3.6)	GAA	25	(5.3)	GGA	12	(2.6)
GTG	7	(1.5)	GCG	2	(0.4)	GAG	6	(1.3)	GGG	5	(1.1)

31	(6.6)	ALA
17	(3.6)	ARG
26	(5.6)	ASN
19	(4.1)	ASP
3	(0.6)	CYS
18	(3.8)	GLN
31	(6.6)	GLU
33	(7.1)	GLY
12	(2.6)	HIS
28	(6.0)	ILE
43	(9.2)	LEU
35	(7.5)	LYS
15	(3.2)	MET
13	(2.8)	PHE
21	(4.5)	PRO
36	(7.7)	SER
41	(8.8)	THR
1	(0.2)	TRP
17	(3.6)	TYR
27	(5.8)	VAL

Poly(A) containing RNA from virus infected cells was electrophoresed on a formaldehyde-agarose gel alongside poly(A) containing RNA from uninfected cells and from mock infected cells treated with actinomycin D. Following electrophoresis, RNA was transferred to nitrocellulose paper as described (31) and hybridized to ^{32}P-labeled HpaII insert of the recombinant plasmid. A single radioactive band of about 1400 bases was visualized only when mRNA from infected cells was used for hybridization (data not shown). The size of the mRNA thus deduced is almost equal to the size of the cloned insert.

An alternative procedure was used to determine whether the entire 5'end sequence was represented in pRSB11. A 78 bp HgiAI/DdeI restriction fragment, downstream from the mRNA start site and 5'end labeled at the HgiAI site was isolated (Fig. 3). This asymmetrically labeled DNA primer was hybridized to poly(A) containing RNA from infected cells and the DNA primer extended on the RNA template using reverse transcriptase as described elsewhere (32 and references therein). A 426 bp HgiAI/HhaI fragment, labeled at the HgiAI site, was used to generate a chemical DNA sequence ladder (lanes G, A, T and C). As illustrated by Fig. 3 (lane 5), the primer was extended 6 nucleotides beyond the cloned RS viral sequence (arrow) implying that this recombinant lacks 6 nucleotides of the 5'end of the mRNA. Extension products seen without the template probably represent artifactual self copying reactions (lane 1). Similar bands are also seen when mRNA was present (lane 5) but are virtually abolished when the reaction is carried out in the presence of actinomycin D (lane 3) which inhibits the double stranded DNA dependent DNA polymerase activity or in the presence of 4 mM sodium pyrophosphate (lane 4) that inhibits the

ribonuclease H activity associated with the reverse transcriptase that degrades the template RNA thereby allowing self priming reactions (33).

DISCUSSION

Sequence analysis of a cDNA plasmid (pRSB11) harboring the RS virus NC protein gene revealed a single long open reading frame encoding 467 amino acids. The other reading frames were extensively blocked throughout thus eliminating the possibility of other proteins encoded within this gene. This is consistent with our previous demonstration (20) of the capsid protein as the sole translation product of RS viral mRNA hybrid selected by pRSB11 plasmid. Amino acid composition of RS NC protein (Table I) shows a relative excess of basic amino acids (Arg, Lys and His) over acidic ones (Glu and Asp). It is relatively rich in proline (21 residues out of 467) but sparse in cysteine (3 residues).

The cDNA insert has 1442 bp of RS viral sequence and lacks 6 nucleotides of the 5'end of the mRNA. Northern blot analysis of mRNA from infected cells revealed a single band of about 1400 nucleotides. Similarly, single discrete poly(A) RNAs, approximately 1140 and 1050 nucleotides in length, reacted with cDNA clones encoding the RS viral M and P genes (unpublished observations). This is in contrast with the reported occurrence of RS viral polycistronic RNAs interpreted by the presence of a single major poly(A) RNA and additional higher molecular weight species reacting with each of the cloned RS viral genes (34). Synthesis of linked messages both *in vivo* and *in vitro* has, however, been reported in the case of VSV (35). We observed such results only when poly(A) RNA was prepared from cells infected at a high multiplicity with a serially passaged virus (data not shown). Such serial undiluted passages can generate defective genomes (36) and transcription of defective genomes lacking intercistronic regulatory regions could yield polycistronic poly(A) RNAs.

Computer analysis of this protein using a homology search program (37) revealed no significant homology with the capsid protein sequences of VSV, influenza virus, coronavirus or tobacco mosaic virus implying that RS virus is evolutionarily distinct. Clustering of basic amino acids within any specific secondary structural domains was not observed when this sequence was analysed by a computer program developed for this purpose (38). This is in contrast to the preferential clustering of basic amino acid residues of the influenza virus polymerase proteins 1 and 3 within alpha helical regions or regions totally devoid of a secondary structure (39). The codon

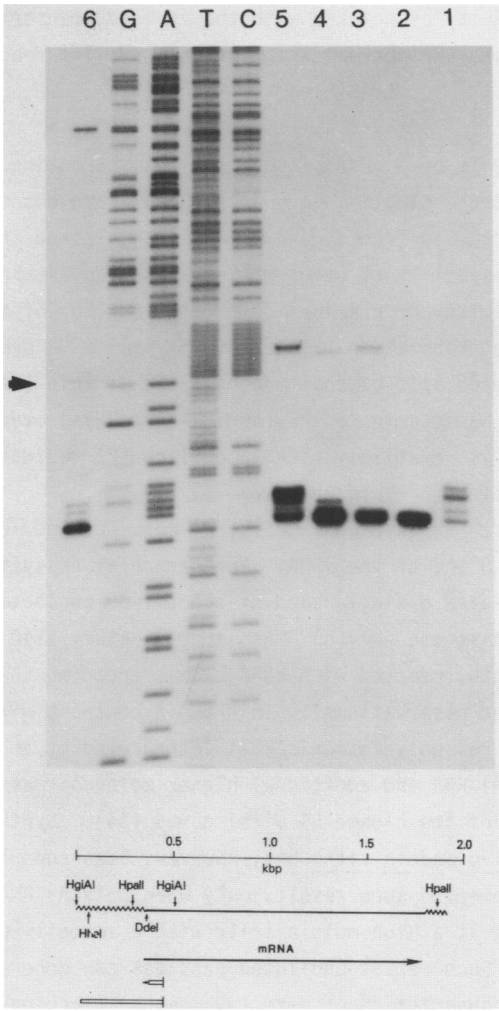

Fig. 3: Size determination of the mRNA encoding the RS viral nucleocapsid protein. The map coordinates of the cloned DNA with respect to the mRNA are schematically illustrated. A 452 bp HgiAI fragment, spanning the 5'end of the cloned gene and adjoining pBR322 sequence (wavyline), was 5'end labeled and digested with HhaI. The HhaI site lies within the pBR322 sequence upstream of the PstI cloning site. The 426 bp HgiAI/HhaI fragment labeled at the HgiAI site was used for chemical DNA sequence analysis. The 78 bp HgiA/DdeI fragment, labeled at the HgiA site is downstream of the mRNA start site and this DNA primer (1 pmole) was denatured and hybridized to poly(A) RNA (6 ug) from infected cells in 40 ul of hybridization buffer containing 80% formamide, 40 mM PIPES (pH 6.2) and 0.4 M NaCl for 16hr at 37°C. Following hybridization, the RNA/DNA hybrid was recovered after two cycles of ethanol precipitation and dissolved in 50 ul of a buffer containing 50 mM Tris-HCl (pH 8.3), 80 mM KCl, 10 mM MgCl2 and 1 mM DTT (complete). Reverse transcriptase (100 units/ml) was added and the

mixture incubated at 37°C for 30 min. Reaction products were recovered following phenol/chloroform extraction and ethanol precipitation and electrophoresed on 8% acrylamide/urea gel. Lanes 5,4 and 3 illustrate the results of a complete reaction or reactions wherein sodium pyrophospate (4 mM) or actinomycin D (48 ug/ml) was present. Reaction products when no RNA was used or when no enzyme was added are also displayed (lanes 1 and 2). The arrow indicates the end of the cloned RS viral sequence. The primer extended product is six nucleotides beyond this position as judged by the number of G:C tails. cDNA extension products with the HpaII insert as the template are illustrated in lane 6. DNA sequence of the HgiAI/HhaI fragment, is shown in lanes G,A,T and C.

usuage for this protein (Table I) showed a remarkable deficiency of CG dinucleotide within the coding region (0.8% versus 3.9% expected on a random basis). Only 5 out of 17 Arg codons are of the CGN type and 3 out of 467 codons are of the type NCG. Such an inherent bias against CG dinucleotide has previously been reported for VSV, influenza virus as well as eukaryotic genomes (40).

Inspection of the 5'end sequence revealed a AXXAUGG sequence flanking the AUG initiator codon. This sequence, wherein the initiator AUG codon is underlined, has been shown to be conserved around functional eukaryotic translational initiaton sites (29). In contrast, there is no counterpart of the canonical eukaryotic polyadenylation signal, AAUAAA, upstream of the 3'poly(A) sequence (28). Comparison of the sequence immediately preceding the poly(A) tail of this gene with the RS virus matrix protein gene and a nonstructural protein gene (unpublished observations) did not reveal any conserved sequence. This is in contrast to the presence of a 3'AUAC sequence in all VSV genes (41) and a 3'AUUC sequence in all Sendai virus genes (42) preceding 7 or 5 U residues that are reiteratively transcribed to yield the poly(A) tail of the mRNAs. Conserved terminal sequences are also present in all the genomic sequences of influenza virus. However, the 3'ends of influenza viral mRNAs do not incorporate the 12-14 nucleotides complementary to the 5'end(s) of the genomic RNAs (43). The poly (A) tails of the mRNAs are, however, generated by reiterative copying of U residues on the template RNAs (44). But there is no sequence homology in the different transcripts upstream of the poly(A) tract. Given these facts, it is tempting to speculate whether the genetic organization of RS virus might resemble influenza virus.

In conclusion, this is the first report of the amino acid sequence of the nucleocapsid protein of RS virus. Experiments are in progress to test whether RS virus NC gene can be expressed *in vivo* using eukaryotic

expression vectors. Such an approach would help determine the specific RNA/protein interactions involved in the nucleocapsid assembly as well as locate the lesions in the several ts mutants (45) by complementation.

ACKNOWLEDGEMENTS

We are grateful to Dr. R.M. Chanock for his unstinting support, constant encouragement and helpful suggestions. Drs. W.J. Lipman and R.J. Feldmann were helpful in asssisting us in the use of the specialized computer programs. Technical support of Ms. Ena Camargo is appreciated.

REFERENCES

1. Chanock, R.M., Kim, H.W., Brandt, C.D. and Parrott, R.H. (1982) in Viral Infections of Humans: Epidemiology and Control. Evans, A.S., Ed., pp. 471-489, Plenum, New York.
2. Huang, Y.T, and Wertz, G.W. (1982) J. Virol. 43, 150-157.
3. Bernstein, J.M. and Hruska, J.F (1981) J. Virol. 38, 278-285.
4. Huang, Y.T. and Wertz, G.W. (1983) J. Virol. 46, 667-672.
5. Banerjee, A.K., Abraham, G. and Colonno, R.J. (1977) J. Gen. Virol. 34, 1-8.
6. Moyer, S.A. and Banerjee, A.K. (1975) Cell 4, 37-43.
7. Kingsbury, D.W. (1977) in Molecular Biology of Animal Viruses. Nayak, D.B. Ed., pp. 367-368, Dekker, New York.
8. Wagner, R.R. (1975) in Comprehensive Virology. Fraenkel-Conrat and Wagner, R.R., Eds., Vol. 4, pp. 1-93, Plenum, New York.
9. Chinchar, V.G. and Portner, A. (1981) Virol. 109, 59-71.
10. Colonno, R.J. and Banerjee, A.K. (1976) Cell 8, 197-204.
11. Glazier, K., Raghow, R. and Kingsbury, D.W. (1977) J. Virol. 21, 863-871.
12. Collins, P.L., Hightower, L.E. and Ball, L.A. (1980) J. Virol. 35, 682-693.
13. Kingsbury, D.W. (1974) Medical Microbiol. Immunol. 160, 73-83.
14. Lazzarini, R.A., Keene, J.D. and Schubert, M. (1981) Cell 26, 145-154.
15. Blumberg, B.M., Leppert, M. and Kolakofsky, D. (1981) Cell 23, 837-845.
16. Leppert, M., Rittenhouse, L., Perrault, J., Summers, D.F. and Kolakofsky, D. (1979) Cell 18, 735-747.
17. Blumberg, B.M. and Kolakofsky, D. (1981) J. Virol. 40, 568-576.
18. Blumberg, B.M., Giorgi, C. and Kolakofsky, D. (1983) Cell 32, 559-567.
19. Joncas, J., Berthiaume, L. and Pavilanis, V. (1969) Virol. 38, 493-496.
20. Venkatesan, S., Elango, N. and Chanock, R.M. (1983) Proc. Natl. Acad. Sci. USA 80, 1280-1284.
21. Ish-Horowicz, D. and Burke, J.F. (1981) Nucl. Acids Res. 9, 2989-2998.
22. Smith, H.O. and Birnstiel, M.L. (1976) Nucl. Acids Res. 3, 2387-2398.
23. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) Cell 15, 1157-1174.
24. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
25. Winberg, G. and Hammarskjold, M.L. (1980) Nucl. Acids Res. 8, 253-264.
26. Queen, C. and Korn, L.J. (1980) Methods in Enzymol. 65, 595-609.
27. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
28. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.

29. Kozak, M. (1981) Nucl. Acids Res. 9, 5233-5252.
30. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schutz, G. (1981) Nucl. Acids Res. 9, 2251-2266.
31. Thomas, S.P. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
32. Venkatesan, S., Baroudy, B. and Moss, B. (1981) Cell 125, 805-813.
33. Myers, J.C. and Spiegelmann, S. (1978) Proc. Natl. Acad. Sci. USA 75, 5329-5333.
34. Collins, P.L. and Wertz, G.W. (1983) Proc. Natl. Acad. Sci. USA 80, 3208-3212.
35. Herman, R.C., Shubert, M., Keene, J.D. and Lazzarini, R.A. (1980) Proc. Natl. Acad. Sci. 77, 4602-4605.
36. Treuhaft, M.W. and Beem, M.O (1982) Infection and Immunity 37, 439-444
37. Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA 80, 726-730.
38. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
39. Sivasubramanian, N. and Nayak, D.P. (1982) J. Virol. 44,321-329.
40. Rose, J.K. and Gallione, C.J. (1981) J. Virol. 39, 519-528.
41. Rose, J.K. (1980) Cell 19, 415-421.
42. Gupta, K.C. and Kingsbury, D.W. (1982) Virol. 120, 518-523.
43. Robertson, J.S. (1979) Nucl. Acids Res. 6, 3745-3757.
44. Robertson, J.S., Schubert, M. and Lazzarini, R.A. (1981) J. Virol. 38, 157-163.
45. Gimenez, H.B. and Pringle, C.R. (1978) J. Virol. 27, 459-464.