Measles Virus Haemagglutinin Gene: Cloning, Complete Nucleotide Sequence Analysis and Expression in COS Cells

By C. GERALD,1,2 R. BUCKLAND,1 R. BARKER,3 G. FREEMAN2 AND T. F. WILD1*

1Unité de Virologie Fondamentale et Appliquée, INSERM U.51, Unité Associée CNRS no. 613, 1 place du Professeur Joseph Renaut, 69371 Lyon Cédex 08, France, 2Laboratory of Immunopathology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts, U.S.A. and 3Plant Breeding Institute, Maris Lane, Trumpington, Cambridge, U.K.

(Accepted 1 August 1986)

SUMMARY

A measles virus (Hallé strain) cDNA library was prepared by cloning virus-induced mRNA directly into the expression vector PCD. Clones corresponding to the measles virus haemagglutinin (HA) gene were isolated and one, PCD-HA-15, which corresponded to the complete mRNA sequence, was further characterized. After transfection into COS-7 cells, measles virus HA antigen was detected by immunofluorescence. The [35S]methionine-labelled HA protein from transfected cells was immunoprecipitated by both polyclonal and monoclonal measles virus antibodies. Analysis by SDS-polyacrylamide gel electrophoresis revealed that the PCD-HA-15 protein migrated in a manner identical to the virus-induced HA. Nucleotide sequence analysis established that the gene contained 1949 nucleotides [exclusive of poly(A)] and coded for a protein containing 617 amino acids. A single hydrophobic domain likely to represent the transmembrane region was identified at the N-terminus. A second overlapping reading frame coded for a protein containing 70 amino acids. This contained a short hydrophobic region (16 amino acids) and had two potential N-glycosylation sites. Comparison of the HA gene of the Hallé strain with the published sequence of the Edmonston strain showed that there was a high degree of conservation (99.3%).

INTRODUCTION

The paramyxoviruses belong to the negative strand RNA virus group (Strauss & Strauss, 1983). They contain a helical nucleocapsid consisting of a 15 kb single-stranded RNA genome (Baczko et al., 1983) with the nucleoprotein (NP), phosphoprotein (P) and large (L) virus proteins. Two transmembrane glycoproteins, the haemagglutinin (HA) and fusion (F) proteins, envelop this complex, whilst a matrix (M) protein lines the inside of the membrane (Choppin & Compans, 1975). Although there is a general replication strategy amongst the paramyxoviruses (Strauss & Strauss, 1983), the arrangement of the genetic material within the genome may vary. Thus, Sendai virus codes for six mRNAs corresponding to the virus structural proteins (Roux & Kolakofsky, 1975), whereas respiratory syncytial virus induces ten mRNAs (Collins et al., 1984). In the latter case, the four additional mRNAs code for proteins which have not been detected in virions and so are designated, non-structural (Collins et al., 1984). In Sendai virus, these proteins may be coded in alternative reading frames within the structural genes (Giorgi et al., 1983). Thus, a virus-induced protein designated C has been attributed to an open reading frame contained in the P gene. The two viruses, Sendai and respiratory syncytial, appear to represent the two extremes of genome organization within the paramyxoviruses. This variation is observed not only at the level of more than one gene product read from a single gene, but also in the organization of the genes in the virus genome (Blumberg et al., 1985; Dickens et al., 1984).
Measles virus is a paramyxovirus which has been extensively studied in recent years because of its clinical importance and its involvement in persistent infections (Rustigian, 1966). Studies on the virus genome structure have shown that the first three genes are organized similarly to those of Sendai virus, that is, 3' leader-NP-P-M (Rozenblatt et al., 1985; Bellini et al., 1985; Richardson et al., 1985). A non-structural protein (C) is coded in an alternative reading frame within the P gene (Bellini et al., 1985). To further these observations and develop tools suitable for the study of their expression, we have attempted to clone the remaining measles virus genes into the expression vector PCD (Okayama & Berg, 1982) via the viral mRNAs. This has enabled us to isolate clones corresponding to the complete mRNA sequence for five of the six measles virus structural genes. In the present paper, we describe the cloning, expression and sequence of the HA gene.

When this work was completed, Alkhatib & Briedis (1986) published the nucleotide sequence of the HA gene of the Edmonston strain of measles virus. This has enabled us to compare the variation in the HA gene of the Hallé and Edmonston strains.

METHODS

Cells and virus. Vero cells were cultivated in Eagle's MEM containing 5% foetal calf serum and antibiotics (100 U/ml penicillin, 100 μg/ml streptomycin). The Hallé strain of measles virus (Horta-Barbosa et al., 1971) was plaque-purified twice in Vero cells; stocks were made from this virus and stored at -70°C.

Preparation of RNA. Vero cells were infected with measles virus at a multiplicity of infection of 1 p.f.u./cell. Twenty-four h later, when half of the cells displayed c.p.e., the total cellular RNA was extracted by addition of 4 M-guanidine thiocyanate, 0.5% sodium N-lauryl sarcosinate, 1 mM-EDTA, 0.0065% anti-foam A (Sigma) and 0.1% 2-mercaptoethanol. The RNA was purified by concentration at 30000 r.p.m. for 24 h at 20°C in the Beckman SW41 rotor through two CsCl cushions of 2.4 M and 5.7 M in 10 mM-EDTA, 25 mM-sodium acetate, pH 5.0. The RNA was dissolved in 0.1 M-NaCl, 5% sodium N-lauryl sarcosinate, 10 mM-Tris, 5% phenol, extracted with phenol/chloroform/isoamyl alcohol (50:49:1, by vol.), and precipitated with 2 vol. ethanol at -20°C. The poly(A)+ RNA was isolated after passage on an oligo(dT)-cellulose column.

Construction of cDNA library. The cDNA library was prepared by a modification of the method of Okayama & Berg (1982). Five μg poly(A)+ mRNA from measles virus-infected cells was denatured by heating at 65°C for 3 min in 15 mM-Tris-HCl pH 7.5 (8 μl) and then quickly cooled on wet ice. The reaction mixture (final total vol. 40 μl) was then adjusted to contain 50 mM-Tris-HCl pH 8.3 (at 42°C), 8 mM-MgCl2, 30 mM-KCl, 0.05 mg/ml actinomycin D, 2 mM-dithiothreitol, 750 μg/ml RNasin, 2 mM each of dATP, dTTP, dGTP, dCTP, 40 μCi [32P]dCTP (800 Ci/mmol) and 2 μg oligo(dT)-tailed plasmid primer PC DVI (Pharmacia) [molar ratio of poly(A)+ mRNA to vector primer DNA approx. 1:15].

The cDNA synthesis was started by the addition of 45 units of avian myeloblastosis virus reverse transcriptase (Life Sciences, St Petersburg, Fla., U.S.A.) and 460 units Moloney murine reverse transcriptase (Bethesda Research Laboratories). The mixture was incubated at 42°C for 30 min. A further 460 units Moloney reverse transcriptase was then added and the incubation continued for a further 30 min at 42°C. All the subsequent steps were carried out according to Okayama & Berg (1982).

The plasmid cDNA was transfected into the MC 1061 strain of Escherichia coli. The bacteria were cultivated in L broth for 1 h. A sample was taken to determine the transformation efficiency. The incubation was continued for a further 10 h at 37°C in the presence of 50 μg/ml ampicillin. The culture was adjusted to 7% DMSO and stored at 1-5 ml aliquots at -70°C.

Identification of measles HA clones. E. coli transformants were screened for the presence of measles virus HA cDNA by in situ colony hybridization (Grunstein & Hogness, 1975). Two-thousand transformants were grown on replica nitrocellulose filter discs, lysed with alkali and hybridized with 32P-labelled nick-translated 340 bp cDNA fragment of the HA gene. This cDNA was kindly provided by Drs B. M. Blumberg and P. C. Dowling, New Jersey Medical School, Newark, N.J., U.S.A., and corresponds to a fragment approximately 600 nucleotides from the 5' end of the mRNA. Colonies that gave positive hybridization signals were grown in L broth containing 50 μg/ml ampicillin and their plasmid DNAs were isolated by standard techniques.

Transfection. Transfection of COS-7 cells was performed by the DEAE-dextran method (Lai & Nathans, 1974). After transfection, cells were incubated in the presence of 150 μM-chloroquine (Luthman & Magnusson, 1983) for 3 h at 37°C.

Immunofluorescence. Cells were fixed with cold acetone 3 days after transfection. Measles virus HA antigen was detected by indirect immunofluorescence with polyclonal or monoclonal anti-HA antibody as previously described (Giraudon & Wild, 1981).

Radiolabelling and immunoprecipitation. Three days after transfection, COS-7 cells were incubated in Eagle's MEM containing 1/10th the normal level of methionine, 5% foetal calf serum and 50 μCi/ml [35S]methionine.
Cloning of measles virus HA

(1021 mCi/mmol, Amersham) for 20 h. Measles virus (Hallé strain) infected COS-7 cells (m.o.i. 0.1 p.f.u./cell) were radiolabelled 24 h after infection for 5 h in the same conditions. Cells were lysed and immunoprecipitated with measles virus-specific polyclonal and monoclonal antibody as previously described (Giraudon et al., 1984). The precipitates were analysed by SDS-PAGE according to Laemmli (1970) and the dried gels were subjected to autoradiography.

Nucleic acid sequencing. Both strands of the virus-specific DNA inserts were sequenced by the method of Maxam & Gilbert (1980).

RESULTS

Construction and identification of cDNA clones

mRNA from measles virus-infected Vero cells was annealed to the oligo(dT)-tailed primer vector of PCD-1. A cDNA was synthesized and the construction completed as described in Methods. The circularized cDNA plasmids were transfected into E. coli MC 1061. Approximately 150,000 transformants were obtained. The clones were subsequently amplified for 1 h at 37 °C in the absence of antibiotics and then 10 h in the presence of 50 µg/ml ampicillin.

Two-thousand clones were screened with a 340 bp fragment representing a sequence of the HA gene which maps 600 nucleotides from the 5' end (B. M. Blumberg & P. C. Dowling, personal communication) and four positive clones were selected for further characterization. The size of the cDNA inserts was measured on agarose gels after digestion with BamHI. Under these conditions, 30 nucleotides at the 5' end and 100 nucleotides at the 3' end of the plasmid DNA will remain attached to the cDNA.

The size of the HA mRNA determined by Northern blotting has been shown to be about 2100 nucleotides (C. Poyer and T. F. Wild, unpublished observations). Two of the clones isolated contained inserts corresponding to the size of the HA mRNA. Restriction enzyme analysis of these clones gave identical results; thus, one (HA-15) was selected for further investigation.

Expression of PCD-HA-15 in COS cells

In the PCD plasmids, the cDNA is inserted between a simian virus 40 (SV40) promoter and polyadenylation signal. The presence of the SV40 origin of DNA replication enables the plasmid to replicate in cells such as COS, which contain a functional SV40 large T antigen (Gluzman, 1981). Transfection of COS cells with the PCD cDNA plasmids will determine whether the cDNA codes for a functional virus protein.

COS-7 cells were transfected with clone PCD-HA-15 by the DEAE-dextran method (Lai & Nathans, 1974). The expression of the measles virus HA antigen was examined on the third day by indirect immunofluorescence using polyclonal and monoclonal antibody (Fig. 1). Up to 5% of the cells expressed virus antigen. The intensity of fluorescence observed in individual cells was equivalent to that found in infected cells. To examine the protein synthesized, the transfected cells were radiolabelled with [35S]methionine and the proteins were immunoprecipitated with measles virus antibody and analysed by SDS-PAGE (Fig. 2). The protein encoded by the plasmid PCD-HA-15 migrated to the same position as the virus-induced HA. It was immunoprecipitated by both monoclonal and virus-specific polyclonal antibody. The total amount of HA synthesized was less in the transfected cells compared to the virus-infected cells. This represents the efficiency of transfection.

Nucleotide sequence of HA

The nucleotide sequence of the measles virus RNA was determined on both strands of the DNA by the method of Maxam & Gilbert (1980) and is shown in Fig. 3.

The first 10 nucleotides AGGGTGCAAG at the 5' end are identical to the consensus sequence at the 5' end of the Sendai virus mRNAs (Gupta & Kingsbury, 1984). In contrast, the polyadenylation site GAATTA6 differs from that of Sendai virus and the other measles virus mRNAs, NP and P (Rozenblatt et al., 1985; Bellini et al., 1985). It is similar, however, to that of the closely related canine distemper virus NP mRNA CATTAn (Rozenblatt et al., 1985). The total length of the HA mRNA excluding the poly(A) is 1949 nucleotides.

The longest open reading frame starts at the ATG at positions 21 to 23 and continues to the
Fig. 1. Transfection of COS-7 cells with PCD-HA-15. Cultures of COS-7 cells were transfected with the measles virus HA plasmid. On the third day, the cultures were fixed and examined by immunofluorescence using a measles virus anti-HA monoclonal antibody.

Fig. 2. Expression of HA proteins in COS-7 cells transfected with cloned measles virus genes. COS-7 cells either infected with measles virus (a to c) or transfected with PCD-HA-15 (d to f) were radiolabelled with $^{35}$S methionine (see Methods). The proteins were immunoprecipitated with measles virus antibody and analysed by SDS-PAGE. (a, d) Anti-HA measles virus monoclonal antibody; (b, e) polyclonal (guinea-pig) measles virus-specific antibody; (c, f) control, no antibody.
TAG and TGA

mRNA would thus contain 76 non-coding nucleotides at the 3' end. The predicted amino acid sequence of the primary translation product of the open reading frame is shown in Fig. 3 and would have a molecular weight of 69,077 excluding carbohydrate residues.

The primary translation product of glycoproteins generally contains a signal sequence at the N-terminus. This initiates the export of the nascent chain across the rough endoplasmic

Fig. 3. Nucleic acid sequence and deduced amino acid sequence of the HA gene of the Hallé strain of measles virus determined from clone PCD-HA-15. The sequence shows the G tail. The numbering starts at the first nucleotide of the HA gene. The amino acid changes in the Hallé sequence compared to PvoHi ~ pro L ~ $ ~ s ~ rArgI ~ a ~ | ~ ea ~ nAr9C ~ vHi$L ~ uMet ~ 1 ~ A ~ pArgPr ~ T ~ V ~ eu~euA ~ Va~uPhe~tPheLeuS ~ r~

TAG termination positions 1872 to 1874. This is followed closely by two further stop signals TAG and TGA at positions 1881 to 1883 and 1884 to 1886 respectively. The measles virus mRNA would thus contain 76 non-coding nucleotides at the 3' end. The predicted amino acid sequence of the primary translation product of the open reading frame is shown in Fig. 3 and would have a molecular weight of 69,077 excluding carbohydrate residues.

The primary translation product of glycoproteins generally contains a signal sequence at the N-terminus. This initiates the export of the nascent chain across the rough endoplasmic
A strong hydrophobic domain near the C-terminus anchors the protein in the plasma membrane (Gupta & Kingsbury, 1984). However, a small number of glycoproteins including the haemagglutinin/neuraminidase (HN) of Sendai virus (Blumberg et al., 1985), simian virus 5 (SV5) (Hiebert et al., 1985), Newcastle disease virus (Schuy et al., 1984) and respiratory syncytial virus (Satake et al., 1985) are anchored by the N-terminus rather than the C-terminus. The mature protein contains the same N-terminus as the primary translation product.

Analysis of the HA structure of measles virus for hydropathy (Kyte & Doolittle, 1982) shows that there is a single hydrophobic domain of 25 amino acids near the N-terminus (amino acids 33 to 58) of the protein (Fig. 4). The C-terminus contains no outstanding hydrophobic domains.

Further examination of the measles virus HA sequence for other open reading frames reveals an ATG at 1727 to 1729 which is preceded by a TATAAA box at positions 1694 to 1699. This is in a reading frame +2 and could code for a protein of 70 amino acids (7645 mol. wt.). The sequence around the ATG contains a G at −3 but not at +4. This would be relatively favourable to initiating translation. The protein is terminated by a TAG termination codon 11 nucleotides before the poly(A). This protein has a small hydrophobic region (16 amino acids) at the N-terminus and two potential N-glycosylation sites, both in hydrophilic regions.

DISCUSSION

Our initial attempts to clone measles virus mRNAs into pBR322 plasmids yielded clones containing DNA complementary only to fragments of the different viral genes. However, by using the cloning method developed by Okayama & Berg (1982), we have now obtained cDNAs corresponding to five or six known viral genes. In each case, the length of the cDNA corresponds to that of the equivalent viral mRNA. For the HA gene described in this paper, we have determined that the 5′ end of the sequence corresponds exactly to the consensus sequence for the start of Sendai virus mRNAs (Blumberg et al., 1985; Gupta & Kingsbury, 1984). This lends added support to the idea that our cDNA represents the full-length viral mRNA sequence and is in agreement with the sequence at the 5′ end determined by Alkhatib & Briedis (1986) for the Edmonston strain.
We have confirmed that our cloned sequences are functionally active by transfecting the PCD-HA plasmids into COS-7 cells. The closure of the plasmids involves insertion of dC:dG base pairs between the SV40 early promoter and the 5' end of the viral message, and such sequences have been shown to reduce transcription efficiency (Gheysen & Fiers, 1982). Our PCD-HA-15 plasmid contains a 35 bp C:G tract, but is nevertheless efficiently expressed in COS-7 cells. The viral antigen may be demonstrated by immunological methods, and the corresponding protein is indistinguishable by PAGE from that obtained by infecting cells with the virus.

The amino acid sequence of the measles virus HA as deduced from the nucleotide analysis of the corresponding gene suggests that the protein may be anchored in the membrane by its amino terminal portion, like the HN proteins of other paramyxoviruses (Blumberg et al., 1985; Hiebert et al., 1985; Schuy et al., 1984) and the neuraminidase of influenza virus (Fields et al., 1981). A stretch of hydrophobic amino acids (35 to 58) could be the transmembrane region as no other sufficiently long hydrophobic domain is observed elsewhere. This region is followed in the C-terminal (exterior) side by a zone containing all five potential N-glycosylation sites predicted by the sequence (amino acids 168 to 238). The HN proteins of Sendai virus and SV5 do possess a potential N-glycosylation site in the equivalent position, but the major concentration (three sites) is found close to the C terminus of the protein. These differences may be related to the neuraminidase activity of these proteins, which is lacking in measles virus HA.

Unlike influenza virus, measles virus is considered to be an antigenically stable entity. A single infection results in life-long protection. The availability of monoclonal antibodies has allowed a more detailed study of this phenomenon. Although some antigenic differences can be detected on the internal components, notably the M protein, the HA and F determinants are extremely stable (Giraudon & Wild, 1981). Comparison of the HA sequence of our Hallé strain, derived from a case of subacute sclerosing panencephalitis, with that of the recently published sequence for the Edmonston strain (Alkhatib & Briedis, 1986) revealed only 13 nucleotide substitutions leading to eight amino acid changes. Analysis of this same strain by W. J. Bellini (personal communication) revealed 14 nucleotide substitutions (10 amino acid changes). However, only seven of the changes (five amino acids) are common to both Edmonston sequences. Thus, it would appear that the difference in the HA of the Hallé strain derived from a case of subacute sclerosing panencephalitis compared to the Edmonston strain is not greater than two separate analyses of the Edmonston strain propagated in different laboratories. Thus, 99.33% of the nucleotides and 98.70% of the amino acids are conserved between the two strains. The correspondence between the HN genes of two strains of Sendai virus, which is also antigenically stable, showed 98.52% of the nucleotides and 97.57% of the amino acids conserved (Blumberg et al., 1985; Miura et al., 1985; Shioda et al., 1986).

The organization of the paramyxovirus genome varies between different members of the family, especially as regards the coding of small, presumably non-structural, peptides. In some cases these are found as distinct separate genes, whereas, in others, they are encoded on a second reading frame within one or more of the major viral genes. Our HA sequence contains a possible open reading frame towards the 3' end of the sequence which could code for a peptide of 70 amino acids. This region of the sequence contains two nucleotide substitutions compared with the Edmonston strain, but neither change the amino acid sequence in the second reading frame. The peptide would contain a short (16 amino acid) hydrophobic region, and two potential N-glycosylation sites are present. No such peptide has yet been identified as a consequence of infection by measles virus.

In this paper, we describe the HA gene of Hallé strain measles virus. The Okayama & Berg (1982) cloning technique has allowed us to isolate cDNAs corresponding to this and four other measles genes. Our library of cloned genes will be used to identify and study the organization and expression of all the measles virus genes.

We would like to thank P. Giraudon for the immunofluorescence studies and Drs P. C. Dowling and B. M. Blumberg for supplying the HA probe. This work was supported by a grant from INSERM (Contrat de Recherche INSERM no. 82.10.57). C.G. is the recipient of a fellowship from the Cancer Research Institute, New York.
REFERENCES


Cloning of measles virus HA


(Received 13 May 1986)