Volume 15 Number 7 1987 Nucleic Acids Research

Nucleotide sequence and *in vitro* expression of rubella virus 24S subgenomic messenger RNA encoding the structural proteins E_1, E_2 and C

David M.Clarke, Tip W.Loo, Ivy Hui, Pele Chong[1] and Shirley Gillam*

Department of Pathology, Faculty of Medicine, University of British Columbia Research Centre, 950 W. 28th Avenue, Vancouver, BC V5Z 4H4, Canada

Received November 4, 1986; Revised and Accepted March 3, 1987

ABSTRACT

The complete nucleotide sequence of the 24S subgenomic mRNA of wild-type M_{33} strain rubella virus has been determined. This RNA is 3,383 nucleotides in length excluding the 3'-terminal poly(A) tract. After the three multiple in-phase termination codons clustered in the 5' terminus of this RNA, there are 81 nucleotides of nontranslated nucleic acid followed by a reading frame of 2,978 nucleotides that encodes the 110 kD precursor of the structural proteins. The 3'-untranslated region is 263 nucleotides. The 110 kD polyprotein is processed to produce nucleocapsid C, the glycoproteins E2 and E1 in that order. Sites of post-translational cleavage to produce E2 and E1 were located using available N-terminal amino acid sequences. RNAs synthesized by transcription in vitro are effective messengers in the rabbit reticulocyte cell-free translation system. Post-translational processing of the structural proteins was observed in the cell-free system supplemented with microsomes from dog pancreas.

INTRODUCTION

Rubella virus is the causative agent of German measles, a common contagious disease of humans associated with a mild fever and general rash. Complications are infrequent but include panencephalitis (1,2) and a transient polyarticular arthritis that is particularly common in adult females (3). The major risk associated with rubella infection occurs during pregnancy when severe damage to the fetus can result including deafness, cataracts, cardiac abnormalities and microcephaly (3,4). Rubella virus has been isolated from peripheral blood lymphocytes of patients with rubella-associated arthritis and adult women immunized with live attenuated rubella virus (5). The elucidation of the molecular basis for rubella replication and persistence is an important goal in rubella research.

The Togaviridae family consists of four genera, Alphavirus, Flavivirus, Pestivirus and Rubivirus (6). Rubella virus, a spherical enveloped virus about 60 nm in diameter (7), is the sole member of the genus Rubivirus (8). It consists of a single molecule of a 40S single stranded RNA of

positive polarity (9). The virus contains three glycosylated, membrane-associated proteins, E1 (Mr = 58,000), E2a (Mr = 47,000) and E2b (Mr = 42,000) and one unglycosylated nucleocapsid protein, C (Mr = 33,000) (10). The two forms of E2 results from heterogeneous glycosylation of an identical apoprotein moiety coded by the E2 gene (10,11). Virus-infected cells contain, in addition to the genomic 40S RNA, a subgenomic 24S mRNA that is derived from the 3' one-third of the genomic 40S RNA (12). The subgenomic 24S mRNA is translated into a precursor to the structural proteins (110 kD) and which is proteolytically processed to give the structural proteins C, E2 and E1 (11, 13). Both 40S and 24S RNAs are capped at their 5'-ends and polyadenylated at their 3'-termini (12). The gene order for rubella virus structural proteins has been shown to be NH_2 - C - E2 - E1 - COOH by sequentially pulse-chase labelling after synchronization of translation initiation with hypertonic salt treatment (13). The partial amino-terminal sequences of E2a, E2b and E1 have been determined (11). It was found that the amino-terminal sequences of E2a and E2b were identical and different from that of E1 (11). The C protein was shown to have a blocked amino-terminal (11).

In this paper we present the nucleotide sequence and describe the *in vitro* transcription and translation of the 24S subgenomic mRNA of M_{33} wild-type rubella virus. The coding regions for the structural genes, C, E2 and E1 have been assigned from the amino-terminal amino acid sequence data (11). An SP_6 *in vitro* transcription system (14) was used to produce biologically active synthetic 24S subgenomic mRNA. A polypeptide of molecular weight 110 kD was immunoprecipitated by rubella virus specific antisera (15) from the *in vitro* translation products (16). Processing of the *in vitro* products was observed when the cell-free system was supplemented with microsomes from dog pancreas endoplasmic reticulum (17). Although the structural proteins of the alphaviruses and rubella virus are synthesized as a precursor polypeptide from the subgenomic mRNA, the processing of the structural proteins of rubella virus appears to differ in some ways from that of the alphaviruses (18).

MATERIALS AND METHODS

Virus and RNA

Rubella virus (M_{33} strain, ATCC VR-315 obtained from Dr. M. Trudel, Virology Research Centre, Institut Armand-Frappier, Quebec, Canada) was plaque purified. The virus stock had an infectivity titre of 10^{6-8}

pfu/ml. Confluent monolayers of Vero cells grown in roller bottles were infected with the stock virus at a multiplicity of 0.1. Infected cells were grown in Eagle's minimum essential medium (Gibco) supplemented with 5% (v/v) heat-inactivated fetal bovine serum. Virus particles were isolated (72 h post-infection) from the medium and purified by sucrose gradient centrifugation (19). Viral RNA was isolated from the purified virus particles by suspending the virus pellet in buffer containing 10 mM Tris-HCl, pH 7.5, 0.1M NaCl, 1 mM EDTA, 0.5% (w/v) SDS and 200 ug/mL proteinase K (BRL) and incubated at 37°C for 1 h. Poly(A)$^+$-RNA was isolated from viral RNA using two cycles of oligo(dT)-cellulose chromatography (20).

Molecular Cloning

Poly(A)$^+$-RNA was transcribed into double-stranded cDNA for cloning. The conditions for first strand and second strand cDNA synthesis were essentially as previously described (21). Construction of a rubella cDNA library in PBR322 vector has been reported (21). In the construction of a λgt-11 library, synthetic oligonucleotides (T_8CT, $G_3A_2TCTAGA$, $GTC_2TCGC_2AT_2GACG_2TA_2GATG_2CAGT_2$) were used as primers in the first strand cDNA synthesis. The synthesized double-stranded cDNA was treated with EcoRI methylase and S-adenosylmethionine (22) to methylate any internal EcoRI cleavage sites. After the ends were repaired with DNA polymerase I (large fragment) EcoRI linkers were ligated onto the ends. Excess linkers were removed from the end by digestion with EcoRI and fractionated on a Bio-Gel A-50M column. Column fractions containing double-stranded cDNA were pooled, ligated in a small volume with EcoRI digested λgt-11 and the ligation mixture was packaged in vitro (23,24). Rubella-specific plaques were selected by using the 5'-P^{32}-labelled synthetic oligonucleotides as probes.

Nucleotides Sequence Analysis

Ten overlapping cDNA's (Fig 2), together representing the entire 24S subgenomic RNA, were excised from plasmid vectors (clones 277 and 48 were digested with PstI and the rest with EcoRI), subcloned into PstI or EcoRI digested M13mp19 and M13mp18, and sequenced by the dideoxynucleotide chain-terminator procedures (25), using synthetic oligodeoxyribonucleotides (S_1-S_{12}, Fig 3) as primers. In some cases, large inserts were shortened by controlled digestion with exonuclease Bal 31 (26) prior to cloning into the M13 vectors.

In Vitro Transcription and Translation

$pSPT_{19}$(C/E2/E1) containing the cDNA insert spanning the entire 24S subgenomic mRNA was constructed by joining the isolated DNA fragments from clones 86 (EcoRI/BstEII), 9(BstEII/StuI) and 277(StuI/PstI), and then inserting the joined 3.3Kb fragment into the $pSPT_{19}$ (Pharmacia), which had been digested with endonucleases Pst I and EcoRI. JC8111 cells were transformed and selected on media containing 50 ug/ml ampicillin. For *in vitro* transcription, plasmids $pSPT_{19}$ (C/E2/E1) and $pSPT_{19}$ (C/E2) were linearized respectively with HindIII and SmaI. Linearized DNA templates were transcribed *in vitro* (27) with SP6 polymerase (Boehringer Mannheim). Translation of RNA transcripts was carried out as described by Pelham and Jackson (28) using cell-free extracts of micrococcal nuclease-treated rabbit reticulocytes (Promega Biotec), in the presence of ^{35}S-methionine at a concentration of 1 uCi per ul of reaction mixture. The translation products were analyzed using 10% SDS-PAGE (29) following immunoprecipitation with antiserum against rubella virus structural proteins (15). $pSPT_{19}$ linearized with HindIII was used as the control. For *in vitro* translation in the presence of microsomes from dog pancreas endoplasmic reticulum (obtained from Dr. P. Walter, Dept. of Biochemistry and Biophysics, University of California, Medical School. San Francisco, California, U.S.A.), 1 ul of microsomes was added to 25 ul of the reaction mixture.

RESULTS

cDNA Cloning

Initially, the first strand of cDNA synthesis was carried out by using $oligo(dT)_{16}$ as a primer, with viral RNA template, reverse transcriptase and four deoxyribonucleoside triphosphates (30). Due to the low concentration of RNA template the yield of cDNA was poor. Another approach was therefore attempted. Since the interaction between oligonucleotides is based on specific base-pairing, the efficiency of priming by an oligonucleotide, such as T_8NN', with RNA containing a poly(A) tail is more specific than that by T_8. The oligodeoxyribonucleotides were used to screen for priming efficiency with the viral RNA by the modified chain termination method (31). It was found that T_8CT gave a specific priming pattern (Fig 1A). The complementary sequence at the 3'-end of viral RNA was determined with T_8CT as a primer in the modified chain termination procedure (Fig 1B). The oligodeoxyribonucleotide, $G_3A_2TCTAGTG$,

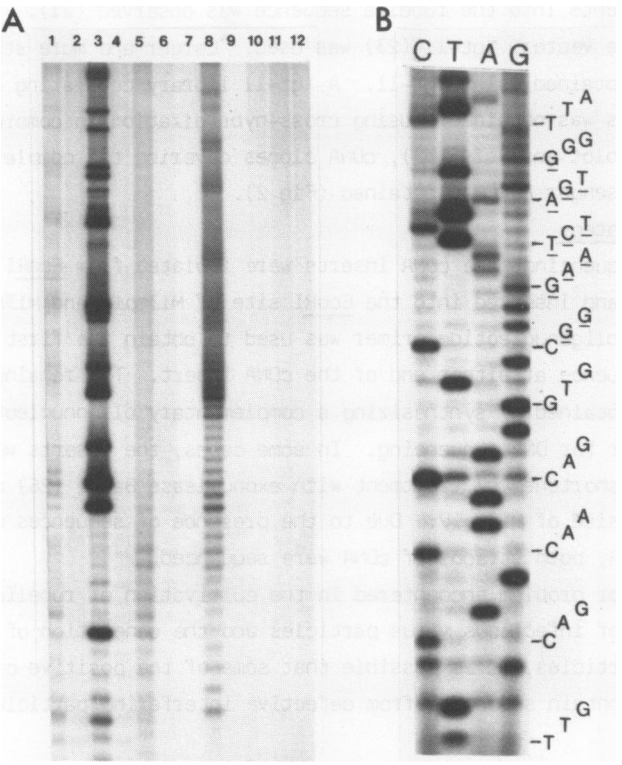


Fig. 1 Complementary sequence of the 3'-end region of viral RNA.
A. Autoradiograph of oligodeoxyribonucleotide primer (T_8NN') screening for specific initiation of cDNA synthesis in rubella viral RNA. The reactions were carried out in the presence of the chain terminator ddTTP (single channel sequencing). The labelled fragments were electrophoresed on a 12% polyacrylamide-7M urea gel. The nucleotide sequence (NN') at the 3'-end of the oligonucleotides are respectively CC, CG, CT, CA, GG, GT, GA, GC, AT, AC, AA and AG from lanes 1 to 12.
B. Complementary sequence of the 3'-end region of rubella viral RNA obtained by using T_8CT as a primer in the modified chain terminator method (31). The underlined sequence corresponds to the oligonucleotide synthesized for cDNA synthesis in subsequent cloning and sequencing.

complementary to the 3'-end region of the viral RNA sequence, was chemically synthesized (32) to serve as a more efficient primer for the synthesis of cDNA as well as a probe for screening recombinant plasmids.

Rubella cDNA libraries have been constructed in pBR322 and pUC_9 vectors using the oligo(dG)/d(C) tailing procedure (30). The cDNA inserts obtained in these plasmids were small and occasionally the integration of

insertion elements into the rubella sequence was observed (21). Therefore a more suitable vector, λgt-11 (23) was used. Larger and more stable inserts were obtained using λgt-11. A λgt-11 library consisting of 20,000 positive clones was obtained. Using cross-hybridization in combination with Northern-blot analysis (33), cDNA clones covering the complete 24S subgenomic messenger RNA was obtained (Fig 2).

Sequencing Strategy

For DNA sequencing, the cDNA inserts were isolated from EcoRI treated λgt-11 clones and inserted into the EcoRI site of M13mp19 and M13mp18. The universal M13 oligonucleotide primer was used to obtain the first 200-300 nucleotide sequence at either end of the cDNA insert. The remainder of the sequence was obtained by synthesizing a complementary oligonucleotide to use as a primer for DNA sequencing. In some cases, the inserts were progressively shortened by treatment with exonuclease Bal31 (26) and cloned into the SmaI site of M13mp19. Due to the presence of sequences of high GC content in cDNA, both strands of cDNA were sequenced.

As the major problem encountered in the cultivation of rubella virus is the low yield of infectious virus particles and the generation of defective interfering particles, it is possible that some of the positive clones obtained may contain sequences from defective interfering particles.

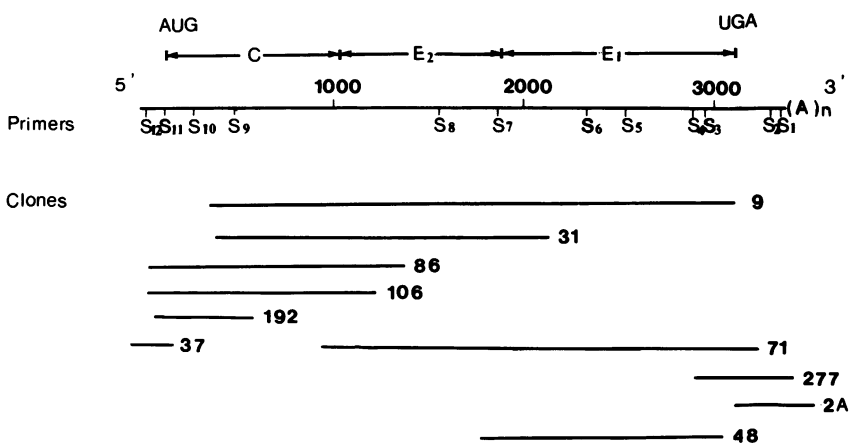


Fig. 2 Sequencing strategy for the 24S subgenomic messenger RNA. The top line shows the coding regions for the structural proteins. The second line indicates the length of this RNA in kilobases. S_1-S_{12} are the complementary synthetic oligonucleotide primers (see Fig. 3 for their sequences). Lines below illustrate the overlapping cDNA clones used for sequence determination.

Therefore, each section of the sequence was determined from at least two different clones. Wherever the sequence differed between two clones (due presumably to heterogeneity in the RNA population or errors introduced during cloning), a third and occasionally a fourth clone was sequenced in this area.

Nucleotide Sequence of the 24S Subgenomic mRNA

Fig. 3 shows the complete nucleotide sequence of the 24S subgenomic mRNA from rubella virus stain M_{33}. The RNA is 3,383 nucleotides in length excluding the 3'-terminal poly(A) tract. The exact length of the poly(A) tail is not known. A poly(A) tract of 60 nucleotides in length was obtained in clone 2A (Fig. 2). A long open reading frame, which spans virtually the entire length of the subgenome beginning from the first AUG (nucleotide 142) is 2,976 nucleotides in length, terminating with a single opal codon (UGA) at nucleotide 3118. This open reading frame could encode a polypeptide of 106,570 daltons. The coding regions for the E2 and E1 proteins have been established from N-terminal amino acid sequences of E2 and E1 (11). Since the C protein has a blocked amino terminus (11), the first AUG codon at nucleotides 142-144 is tentatively assigned to initiate the synthesis of the capsid protein.

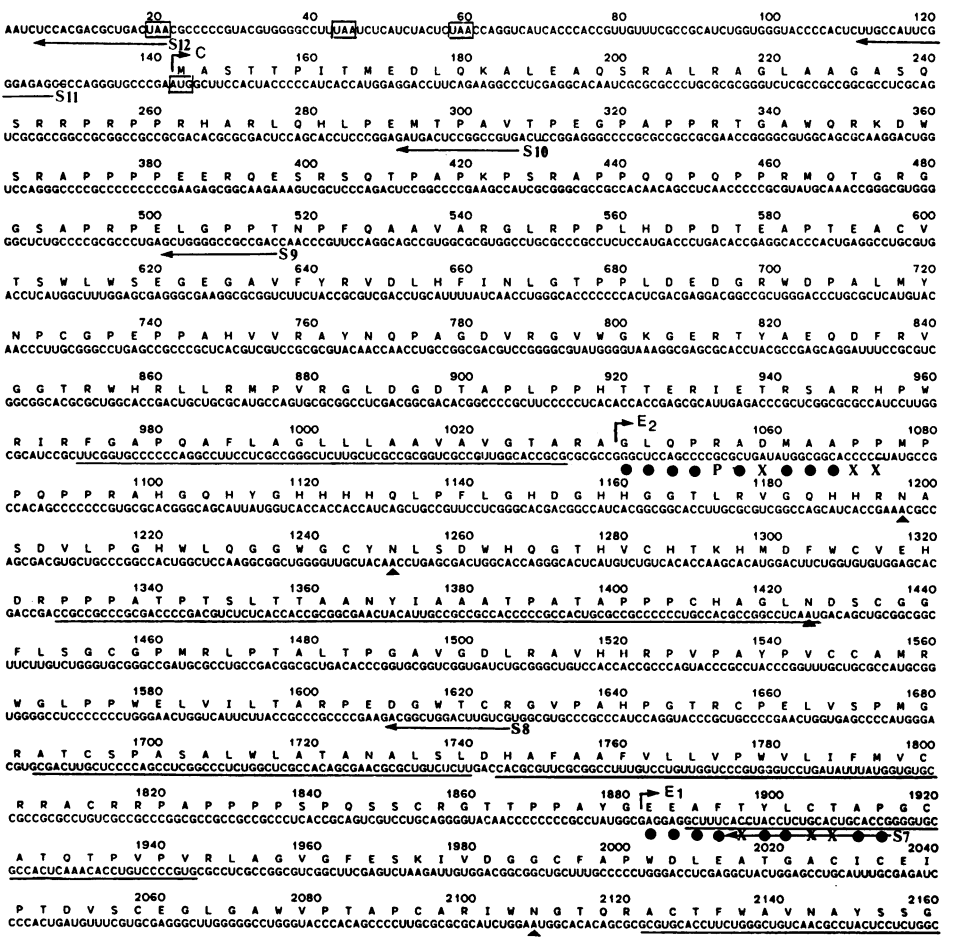
There are three in-phase termination codons clustered in the 5'-noncoding region of the 24S RNA. A similar arrangement is found in other alphaviruses (34). The 5'-terminal cap site for the 24S subgenomic RNA is not known. We have sequenced at least five different cDNA clones in the 3' region of the 24 S RNA. Three in phase stop codons are observed in all 5 cDNA clones. The 3' nontranslated region from the putative termination codon (nucleotides 3118-3120) is 263 nucleotides long, which is similar to that of Semliki Forest virus (18). It is not known if the other two downstream UAG stop codons near the 3' poly A tract are used for termination of the translation of the structural proteins.

The codon usage in the structural region is similar to that found in the structural protein region of alphaviruses (18,35). In many cases codons containing G or C are preferred (Table I), which probably reflects the relative stability of the codon-anticodon pairing. The base analysis of the coding region for the structural proteins is 15% A, 30% G, 39% C and 16% U. The GC content is 69%.

Characteristics of the Structural Proteins

The deduced amino acid sequence for the structural protein is shown in Fig. 3. Table II gives the amino acid compositions of the structural

proteins. The C protein has 299 amino acids and is rich in proline and arginine residues. The clusters of proline and basic amino acids in the C region may be involved in binding to the 40S genomic RNA in the viral nucleocapsid. Near the carboxy terminal of C is a stretch of 21 uncharged residues. The E2 protein has 281 amino acids, rich in proline and cysteine, and contains three potential asparagine-linked glycosylation sites in the form of Asn-X-Ser (36). Three long stretches of hydrophobic regions are observed in the E2 coding region. The E1 protein has 412 amino acids, and is also rich in proline and cysteine. It contains three possible asparagine-linked glycosylation sequences of the form Asn-X-Thr



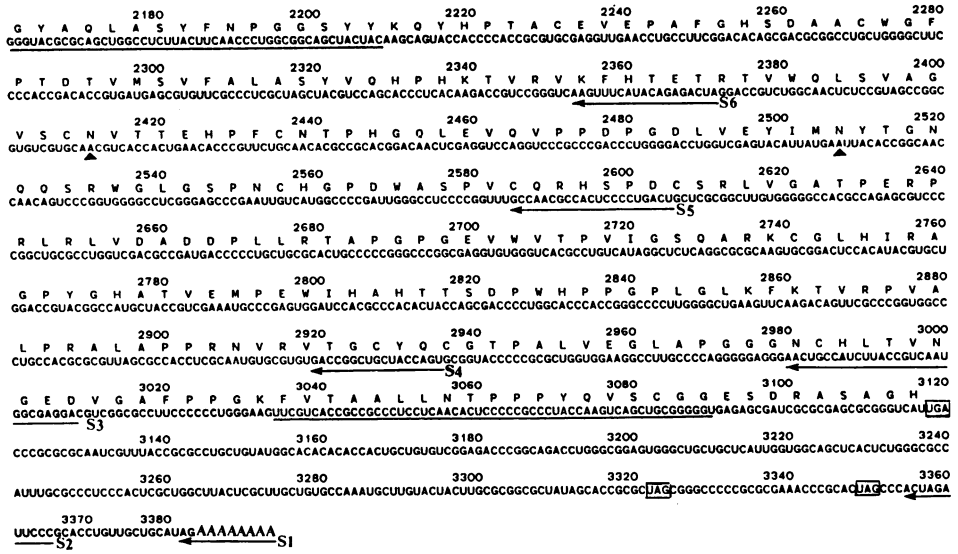


Fig. 3 Nucleotide sequence of the 24S subgenomic mRNA. The complete translated sequence is shown. The amino-terminal ends of each protein are indicated by arrows, potential asparagine-linked glycosylation sites are shown by triangles, and the hydrophobic regions of the viral proteins are underlined. The initiation and termination codons are boxed. The reported amino terminus of E2 and E1 are denoted by solid circles (X refers to an undetermined amino acid, P. Proline was reported (11).

Table I. Codon Usage in the Structural Region [Nucleotides 142-3117]

Phe	UUU	5	Ser	UCU	8	Tyr	UAU	2	Cys	UGU	5
	UUC	20		UCC	10		UAC	20		UGC	31
Leu	UUA	1		UCA	2	Och	UAA	0	Opal	UGA	0
	UUG	6		UCG	10	Amb	UAG	0	Trp	UGG	28
Leu	CUU	8	Pro	CCU	21	His	CAU	14	Arg	CGU	9
	CUC	29		CCC	50		CAC	32		CGC	42
	CUA	0		CCA	13	Gln	CAA	13		CGA	56
	CUG	27		CCG	45		CAG	24		CGG	11
Ile	AUU	6	Thr	ACU	18	Asn	AAU	6	Ser	AGU	1
	AUC	6		ACC	37		AAC	15		AGC	14
	AUA	3		ACA	8	Lys	AAA	1	Arg	AGA	0
Met	AUG	16		ACG	7		AAG	12		AGG	3
Val	GUU	6	Ala	GCU	10	Asp	GAU	7	Gly	GGU	12
	GUC	34		GCC	58		GAC	30		GGC	53
	GUA	4		GCA	5	Glu	GAA	10		GGA	7
	GUG	20		GCG	37		GAG	31		GGG	23

Table II. Amino Acid Composition *of Rubella Structural Proteins

Amino Acid	C	E_2	E_1
Alanine	12.4	11.7	9.7
Arginine	11.4	6.4	4.6
Asparagine	1.3	1.8	2.9
Aspartate	4.0	3.6	3.6
Cysteine	0.7	5.0	4.9
Glutamine	5.0	2.8	3.4
Glutamate	6.0	1.4	4.6
Glycine	8.4	9.6	10.4
Histidine	2.7	7.8	3.9
Isoleucine	1.3	1.1	1.9
Leucine	7.4	8.5	6.1
Lysine	1.3	0.4	1.9
Methionine	2.0	2.5	0.7
Phenylalanine	2.0	2.1	3.2
Proline	15.1	14.2	10.7
Serine	4.0	4.3	5.1
Threonine	6.7	6.4	7.8
Tryptophan	2.7	3.2	2.7
Tyrosine	1.3	1.8	3.2
Valine	4.3	5.3	8.7
Total	100.0	100.0	100.0

*Expressed as mole percentages of amino acids.

(36). Three long contiguous sequences of uncharged hydrophobic amino acids are also present in the E1 coding region.

Since the E2 and E1 proteins are integral membrane proteins, the long stretches of uncharged hydrophobic regions which are close to the amino- and carboxyl-terminus of E2 and E1 are likely to function as either the signal peptides for translocation or the membrane-spanning domains for protein anchoring. The long stretch of the hydrophobic region close to the carboxy terminal of the capsid protein may contain the signal sequence for the translocation of E2.

Expression of cDNA Transcripts in vitro

The 24S subgenomic RNA codes for all of the structural proteins of rubella virus and serves as messenger RNA for these proteins both *in vivo* (13) and *in vitro* (12). In cell free lysates programmed with 24S RNA isolated from rubella-infected cells, the only translation product is a 110 kD protein, which is the precursor to all of the structural proteins. Thus, the cDNA transcripts of the 24S RNA will direct the *in vitro* synthesis of a 110 kD polypeptide, immunoprecipitable by antiserum against rubella structural proteins. The constructed transcription plasmid

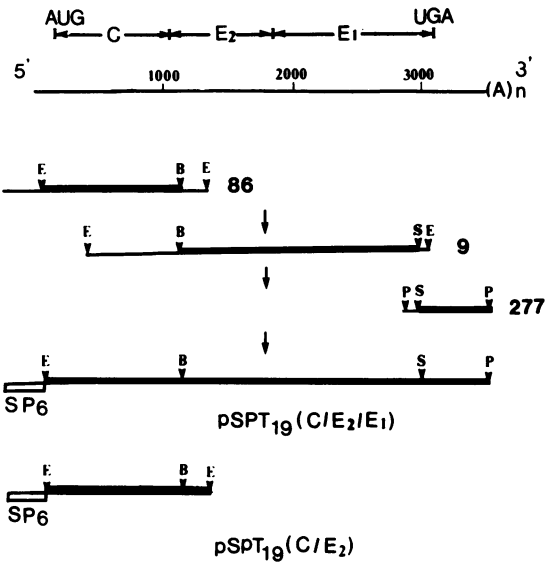


Fig. 4 cDNA of transcription plasmids $pSPT_{19}$(C/E2/E1) and $pSPT_{19}$(C/E2). Plasmid $pSPT_{19}$ (C/E2/E1) was linearized at the HindIII site and plasmid $pSPT_{19}$(C/E2) at the SmaI site. RNA was synthesized by run-off transcription from the SP_6 promoter with SP_6 RNA polymerase. The top line shows the coding regions for the structural proteins. The second line indicates the scale in kilobases of the 24S subgenomic mRNA. Transcription plasmids were constructed as described in Methods. Transcribed regions are shown by thick lines. E, EcoRI; B, BamHI; S, StuI; P, PstI;l Sp_6; Sp_6 prometer.

($pSPT_{19}$(C/E2/E1) containing the cDNA of the entire 24S RNA (Fig 4) was linearized at the HindIII site and transcripts synthesized in vitro with SP_6 RNA polymerase (37). The transcripts were translated in a cell-free system of micrococcal nuclease-treated rabbit reticulocytes (28) and the immunoprecipitated products were identified by electrophoresis in SDS-polyacrylamide gels (Fig 5A).

It was found that cDNA transcripts from $pSPT_{19}$ (C/E2/E1) had the same messenger activity as did the 24S subgenomic RNA in vitro. A polypeptide of molecular weight of 110 kD was immunoprecipitated from the translation products by antisera against rubella virus structural proteins (Fig 5A, lane 5). A polypeptide of 97 kD was observed (data not shown) when the transcription plasmid containing the coding regions of C, E2 and E1 lacking the carboxy terminal sequence of E1 and the 3'-noncoding region (nucleotides 95-3010). Thus, the termination codon at nucleotides

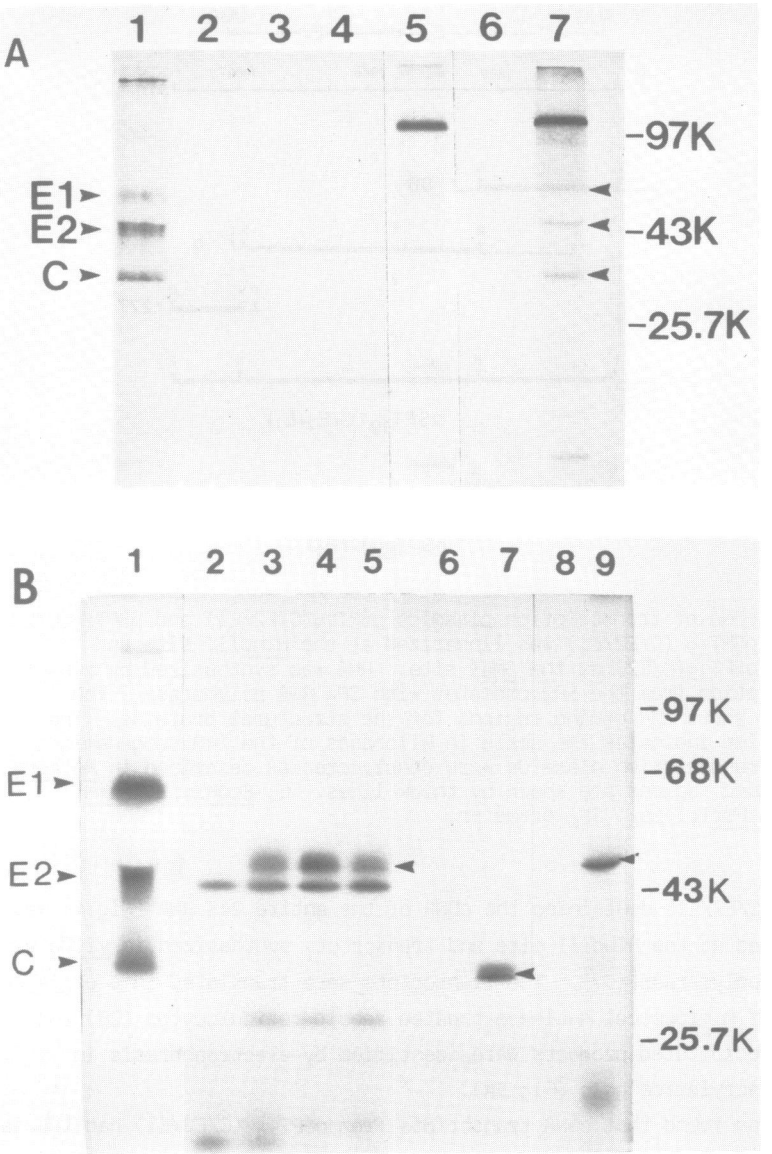


Fig. 5 Polypeptides directed by the synthetic transcripts in the rabbit reticulocyte cell-free translation system.
Products labelled with [S^{35}]-methionine were immunoprecipitated with rubella virus-specific antisera and separated by electrophoresis in a 10% SDS-polyacrylamide gel and detected by autoradiography. The positions of the molecular weight markers are shown. E1, E2 and C refer to the structural proteins of rubella virus. $pSPT_{19}$ linearized at the HindIII site was used as a control.

A. Transcripts from $pSPT_{19}$(C/E2/El).
Immunoprecipitated translation products of transcripts prepared from pSPT T_{19}(C/E2/El). Lanes 2, 4 and 6 show the products of pre-immune sera and lanes 3, 5 and 7 show the products using rubella virus specific antisera. Lane 2 and 3 are products of $pSPT_{19}$. Lanes 4 and 5 are products of $pSPT_{19}$(C/E2/El). Lanes 6 and 7 are products of $pSPT_{19}$(/E2/El) in the presence of microsomes. Arrows indicate the positions of El, E2 and C proteins. Lane 1 is the immunoprecipitated [S^{35}]-methionine labelled rubella virus.
B. Transcripts from $pSPT_{19}$ (C/E2).
Total translation products in the absence of cap, $pSPT_{19}$ (lane 2) $pSPT_{19}$ (C/E2)(lane 3). In the presence of m^7GpppG, $pSPT_{19}$(C/E2)(lane 4) and m^7GpppA,$pSPT_{19}$(C/E2)(lane 5). Products immunoprecipitated by antibody against rubella capsid protein, $pSPT_{19}$ (lane 8), $pSPT_{19}$(C/E2)(lane 9). [S^{35}]-labelled rubella virus was immunoprecipitated by antibody against rubella capsid protein (lane 7) and by pre-immune sera (lane 6).

3118-3120 was used to terminate translation of the precursor protein and the long open reading frame in the 24S subgenomic RNA to encode a polypeptide of 110 kD.

Another transcription plasmid which contained C and part of E2 coding regions ($pSPT_{19}$(C/E2), Fig 4) was constructed and the synthesized transcripts were translated in vitro. Similarly, no free capsid protein was immunoprecipitated by antisera against nucleocapsid protein from the translation products and a polypeptide with a molecular weight of 46 kD was immunoprecipitated (Fig 5B, lane 9). This polypeptide corresponds to the coding capacity of the cDNA insert which encoded 423 amino acid residues. Uncapped transcripts were translated as efficiently as capped transcripts (Fig 5B, lanes 4-5).

In a cell-free system, supplemented with microsomes from dog pancreas endoplasmic reticulum (17), correct processing of the in vitro products occurred (18). This system reproduced the protein synthesis, proteolytic cleavages, membrane protein segregation, and glycosylation observed in infected cells. We have carried out the in vitro translation of cDNA transcripts synthesized from plasmid $pSPT_{19}(C/E_2/E_1)$ with rabbit reticulocytes in the presence of microsomes from dog pancreas endoplasmic reticulum. It was found that correct processing of the in vitro products occurred as is evident from the migration of the immunoprecipitated proteins (Fig 5A, lane 7), which is identical to that of the corresponding proteins in the control sample (Fig 5A, lane 1). It is clear that translocation of E2 and El takes place. The low yield of E2 and El observed in the immunoprecipitation could be either due to the specificity

of the anitsera used, or the incomplete glycosylation of E2 and E1, as incomplete glycosylation in vitro is not uncommon (38, 39). It is possible that the carbohydrate moieties are likely to be very important in maintaining the antigenic properties of E2 and E1.

DISCUSSION

We have cloned and sequenced the 24S subgenomic mRNA of rubella virus M_{33} strain. The RNA contains a single long open reading frame of 2,978 nucleotides that encodes the 110 kD precursor of the structural proteins. The first AUG codon at nucleotide 142-144 is tentatively assigned to initiate the synthesis of 110 kD precursor. However, according to the analysis of sequences upstream from the translational start site in eukaryotic mRNAs, the sequence $CC^{A}_{G}CCAUG(G)$ emerges as a consensus sequence for eukaryotic initiation sites (40). Therefore the second AUG codon at nucleotides 166-168 (UCACCAUGG) appears to be a better candidate for the AUG initiator codon. Another argument in favour of the second AUG codon as the initiator codon is that the first AUG codon is within the inverted repeat sequence (GCCAUUCGGG, nucleotide 114-122) and can base-pair with UAC in a hairpin configuration. It is possible that the replicase/transcriptase may not be able to initiate protein synthesis in such a configuration. In vitro site-specific mutagenesis by using synthetic oligonucleotides as mutagens (41) can be used to define the initiator codon for the C protein. This work is currently in progress.

The rubella genome consists of a single positive-strand RNA molecule that is about 11,000 nucleotides long (12). This genomic RNA serves as the mRNA for the nonstructural proteins that probably function as the replicase/transciptase activities required for replication of the viral RNA and for the transcription of the 24S subgenomic mRNA encoding the virion structural proteins. The 24S RNA contains sequences identical to the 3'-terminal one-third of the genomic 40S RNA. The nucleotide sequences of the genomic RNA preceding and including the beginning of the 24S RNA may contain signals important in the termination of translation of the nonstructural polyprotein precursor as well as sequences important for the transciption of 24S subgenomic RNA. Three in-phase termination codons clustered at the 5'-untranslated region of the 24S RNA were found. These multiple stop codons are used to insure the complete termination of the translation of the non-structural proteins. Interestingly, three in-phase termination codons are also observed in the alphaviruses at the

5'-untranslated region of the 26S subgenomic RNA (34). An inverted repeat (GCCAUUCGGG) at nucleotides 114-122 and nucleotides 137-146 that can be arranged in a hairpin configuration may function as a regulatory signal to allow the synthesis of 40S genomic RNA and 24S subgenomic mRNA to be regulated independently from each other. From our in vitro translation studies, the opal termination codon at nucleotides 3118-3120 is used for the termination of the translation of the 110 kD precursor (Fig 5A). The biological significance of the two in-phase stop codons (at nucleotides 3322-3324 and 3349-3351) near the 3'-terminus before the poly(A) tract is unknown at present.

Rubella virus contains three major structural proteins, C, E2 and E1 (10). The E2 glycoprotein in mature virus exists in multiple forms that are separable on the basis of molecular weight differences on SDS-polyacrylamide gel electrophoresis. The heterogeneity in E2 is due to post-translational modification of a single gene product E2 (10). From the deduced amino acid sequence of E2 and E1, three potential asparagine-linked glycosylation sites (36) are observed both in E2 and E1 coding regions. It is not known which of these are glycosiated in E2a, E2b and E1. Carbohydrate units of the viral membrane proteins may function as either the sorting signals necessary to route these proteins from the rough endoplasmic reticulum to the plasma membrane; to stabilize a certain conformation that can be transported; or to maintain the correct conformation necessary for their antigenecity. Therefore, it is necessary to study the carbonydrate processing in E2 and E1 in order to understand the biological function of E2a, E2b and E1.

Very little is known about the processing of the rubella virus structural proteins. It is known that translation is initiated at a single site on the subgenomic 24S mRNA and the proteins are synthesized as a precursor in the order capsid-E2-E1. The capsid protein does not seem to act as an autoprotease responsible for the first cleavage of the nascent chain in the cytoplasm as is observed in other alphaviruses (18)(Fig 5). The in vitro transcription from plasmids in the presence of microsomes strongly suggests that the capsid protein, E2 and E1 are cleaved by the signal peptidase on the luminal surface of endoplasmic reticulum. The long stretches of the hydrophobic amino acid residues present in the glycoproteins, E2 and E1 may function either as signal peptides or membrane-spanning domains in the translocation of E2 and E1. The identification of the signal sequences and the membrane-spanning domains

for E2 and E1 is currently under investigation by *in vitro* site-specific mutagenesis with synthetic oligodeoxyribonucleotides as mutagens (41).

Analysis of rubella 24S subgenomic mRNA sequence has provided a fuller understanding of the coding regions of the genes for the structural proteins and their antigenic determinants. The rubella virus sequence presented in Fig. 3 was compared to the sequences of the structural proteins of the other alphaviruses (18,35). No homology was found at either the amino acid or nucleotide level. The overall lack of nucleotide and amino acid homology between rubella virus and the alphaviruses confirms the classification of rubella virus in its own genus in the Togaviridae family.

ACKNOWLEDGMENT

This work was supported jointly by grants from the Medical Research Council of Canada, British Columbia Health Care Research Foundation, Alberta Heritage Foundation for Medical Research and Health and Welfare Canada. P.C. is an MRC Postdoctoral Fellow. T.L. is a B.C. Children's Research Foundation Postdoctoral Fellow.

[1]Present address: Connaught Research Institute, 1755 Steeles Avenue West, Willowdale, Ontario M2R 3T4, Canada

*To whom correspondence should be addressed

REFERENCES

1. Townsend, J.J., Baringer, J.R., Wolinsky, J.S., Malamud, N., Mednick, J.P., Panitch, H.S., Scott, R.A., Oshiro, L.S. and Cremer, N.E. (1975). New Eng. J. Med. 292, 994-998.
2. Cremer, N.E., Oshiro, L.S., Weil, M.L., Lennette, E.H., Habashi, H.H. and Carney, L. (1975). J. Gen. Virol. 29, 143-153.
3. Yanez, J.E., Thompson, G.R., Mikkelson, W.M. and Bartholomew, L.E. (1966). Annals. Int. Med. 64, 772-777.
4. Menser, M.A. and Reye, R.D.K. (1974). Pathology 6, 215-222.
5. Chantler, J.K., Ford, D.K. and Tingle, A.J. (1982). Lancet 1, 1323-1325.
6. Brinton, M.A. (1980). In Schlesinger, R.W. (ed.). The Togaviruses. Academic Press, Inc. New York. pp 623-666.
7. Holmes, I.H., Wark, M.C. and Warburton, M.F. (1969). Virology 37, 15-25.
8. Melnick, J.L. (1976). Prog. Med. Virol. 22, 211-221.
9. Hovi, T. and Vaheri, A. (1970). Virology 42, 1-8.
10. Oker-Blom, C., Kalkkinen, N., Kääriäinen, L. and Petterson, R.F. (1983). J. Virol. 46, 964-973.
11. Kalkkinen, N., Oker-Blom, C. and Petterson, R.F. (1984). J. Gen. Virol. 65, 1549-1557.
12. Oker-Blom, C., Ulmanel, I., Kääriäinen and Petterson, R.F. (1984). J. Virol. 49, 403-408.

13. Oker-Blom, C. (1984). J. Virol. 51, 806-812.
14. Butler, E.T. and Chamberlin, M.J. (1982). J. Biol. Chem. 257, 5772-5778.
15. Loo, T.W., MacDonald, I., Clarke, D.M., Trudel, M., Tingle, A. and Gillam, S. (1986). J. Virol. Meth. 13, 149-159.
16. Pelham, R.B. and Jackson, R.T. (1976). Eur. J. Biochem. 67, 247-256.
17. Lingappa, V.R., Katz, F.N., Lodish, H.F. and Blobel, G. (1978). J. Biol. Chem. 253;8667-8670.
18. Garoff, H., Kondor-Koch, C. and Riedel, H. (1982). Current Topics Microbiol. Immunol. 99, 1-50.
19. Trudel, M. and Payment, P. (1980). Can. J. Microbiol. 26, 1334-1339.
20. Aviv, H. and Leder, P. (1972). Proc. Natl. Acad. Sci. U.S.A. 69, 1408-1412.
21. Chong, P., Hui, I., Loo, T. and Gillam, S. (1985). FEBS Letters, 192, 47-52.
22. Greene, P.J., Poonian, M.S., Nussbaum, A.L., Tobias, L., Garfin, D.E., Boyer, H.W. and Goodman, H.M. (1975). J. Mol. Biol. 99, 237-261.
23. Young, R.A. and Davis, R.W. (1983). Proc. Natl. Acad. Sci. U.S.A. 50, 1194-1198.
24. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, D. Quon, Sim, D.K. and Efstratiadis, A. (1978). Cell 15, 687-701.
25. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
26. Lau, P.P. and Gray, H.B. (1979). Nucl. Acids Res. 6, 331-357.
27. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T. and Green, M.R. (1984). Nucleic Acids Res. 12, 7035-7056.
28. Pelham, R.B. and Jackson, R.T. (1976). Eur. J. Biochem. 67, 247-256.
29. Laemmli, U.K. (1970). Nature 227, 680-685.
30. Efstratiadis, A. and Villa-Komaroff, L. (1979). In Setlow, J.K. and Hollaender, A. (eds). Genetic engineering, Plenum Press, New York, Vol. 1, pp 15-36.
31. Boss, J.M., Gillam, S., Zitamer, R.S. and Smith, M. (1982). J. Biol. Chem. 256, 12958-12961.
32. Atkinson, T. and Smith, M. (1984). In Gait, M.J. (ed). Oligonucleotide Synthesis. IRL press, pp 83-114.
33. McMaster, G.K. and Carmichael, G.G. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 4835-4838.
34. Ou, J-H, Rice, C.M., Dalgarno, L., Strauss, E.G., and Strauss, J.H. (1982). Proc. Natl. Acad. Sci. U.S.A. 79, 5235-5239.
35. Rice, C.M. and Strauss, J.H. (1981). J. Mol. Biol. 150, 315-340.
36. Neuberger, A., Gottschalk, A., Marshall, R.D. and Spiro, R.G. (1972). In Gottschalk A. (ed). The Glycoproteins, Elsevier Publishing, Amsterdam, pp 450-490.
37. Contreras, R., Cheroutre, H., Degrave, W. and Fiers, W. (1982). Nucleic Acids Res. 10, 6353-6361.
38. Perara, E. and Lingappa, V.R. (1985). J. Cell. Biol. 101, 2292-2301.
39. Spiess, M. and Lodish, H.F. (1986). Cell. 44, 177-185.
40. Kozak, M. (1984). Nucl. Acids Res. 12, 857-872.
41. Zoller, M. and Smith, M. (1982). Nucl. Acids Res. 10, 6487-6500.