Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

VIROLOGY 184, 531–544 (1991)

A Polycistronic mRNA Specified by the Coronavirus Infectious Bronchitis Virus

D. X. LIU,* D. CAVANAGH,† P. GREEN,† AND S. C. INGLIS*,‡

*Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, United Kingdom; ‡Immunology Limited, 184 Cambridge Science Park, Milton Road, Cambridge CB4 4GN, United Kingdom; and †AFRC Institute of Animal Health, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA, United Kingdom

Received April 19, 1991; accepted May 24, 1991

The third largest of the nested set of subgenomic mRNAs (mRNA3) from the coronavirus infectious bronchitis virus (IBV) contains three separate open reading frames (3a, 3b, and 3c) which are not present on the next smallest of the mRNAs, suggesting that this mRNA may be functionally polycistronic. However, although a protein product has been identified from the 3c open reading frame, to date the coding function of 3a and 3b has not been established. We present nucleotide sequence data suggesting that each of the three open reading frames is conserved in a variety of different IBV strains and further show, through the preparation of monospecific antisera against bacterial fusion proteins, that IBV-infected cells contain small amounts of the products of these ORFs. *In vitro* translation studies using synthetic mRNAs containing the 3a, 3b, and 3c open reading frames suggest strongly that all three proteins can be translated from a single molecular species, and expression studies carried out in intact cells support this conclusion. Thus mRNA3 of IBV appears to be functionally tricistronic.

INTRODUCTION

Avian infectious bronchitis virus (IBV) is the type species of the Coronaviridae, a family of viruses which possess a large continuous positive-sense single-stranded RNA genome. The genomic RNA of these viruses is capped and polyadenylated and is by itself infectious (Schochetman *et al.*, 1977; Lomniczi, 1977), indicating that it can act directly in the infected cell as a virus mRNA. Boursnell *et al.* (1987), through cloning of cDNA, deduced the complete nucleotide sequence of the genomic RNA of IBV and showed it to be 27 kb in length, with at least 10 distinct open reading frames (ORFs). However, it appears that most of these are not translated from genomic RNA but are expressed from a set of subgenomic mRNAs. Available evidence indicates that in virus-infected cells, six subgenomic mRNA species are produced. These mRNAs, designated 2 to 6 (the genomic mRNA is designated mRNA1) (Cavanagh *et al.*, 1990), range in length from about 2 to 7 kb and have been shown to share a common 3′ terminus and to form a "nested" set structure (Stern and Kennedy, 1980a, b). Three of these, mRNAs 6, 4, and 2, appear to be functionally monocistronic; *in vitro* translation studies have shown that they encode the major virion structural proteins nucleocapsid (N—mRNA6), membrane (M—mRNA4), and spike (S—mRNA2), and in each case the expressed ORF is the only one which is not present on the next smallest mRNA in the nested set (Fig. 1). Each of the other mRNAs, however, contains more than one ORF in its "5′ unique" region, suggesting that it may be functionally polycistronic. We describe here studies designed to test this possibility for mRNA3.

Nucleotide sequence analysis has shown that the 5′ terminal sequence of mRNA3 contains three ORFs, designated 3a, 3b, and 3c, which are not represented on mRNA4 and which have the potential to encode three polypeptides of molecular weights 6.7K, 7.4K, and 12.4K, respectively (Boursnell *et al.*, 1985). We have previously shown that one of these, 3c, is indeed expressed in IBV-infected cells (Smith *et al.*, 1990), but so far the coding function of 3a and 3b has not been definitely established. We report that infected cells also contain proteins corresponding to the 3a and 3b ORFs, indicating that mRNA3 is indeed functionally polycistronic. Furthermore translation studies using artificial mRNAs synthesized by *in vitro* transcription indicate that all three ORFs may be expressed from a single species of mRNA, suggesting that mRNA3 of IBV is tricistronic.

MATERIALS AND METHODS

Virus and cells

The Beaudette strain of IBV was grown in the allantoic cavity of 11-day-old embryonated chicken eggs, obtained from specified pathogen-free flocks; the virus was assayed by plaquing on chick kidney (CK) cell cultures prepared from 2- to 3-week-old birds. Additional IBV strains were grown in eggs to produce genomic RNA for nucleotide sequencing. These strains were: USA/M41/41 (isolated in 1941); UK/183/66 (Dawson

0042-6822/91 $3.00

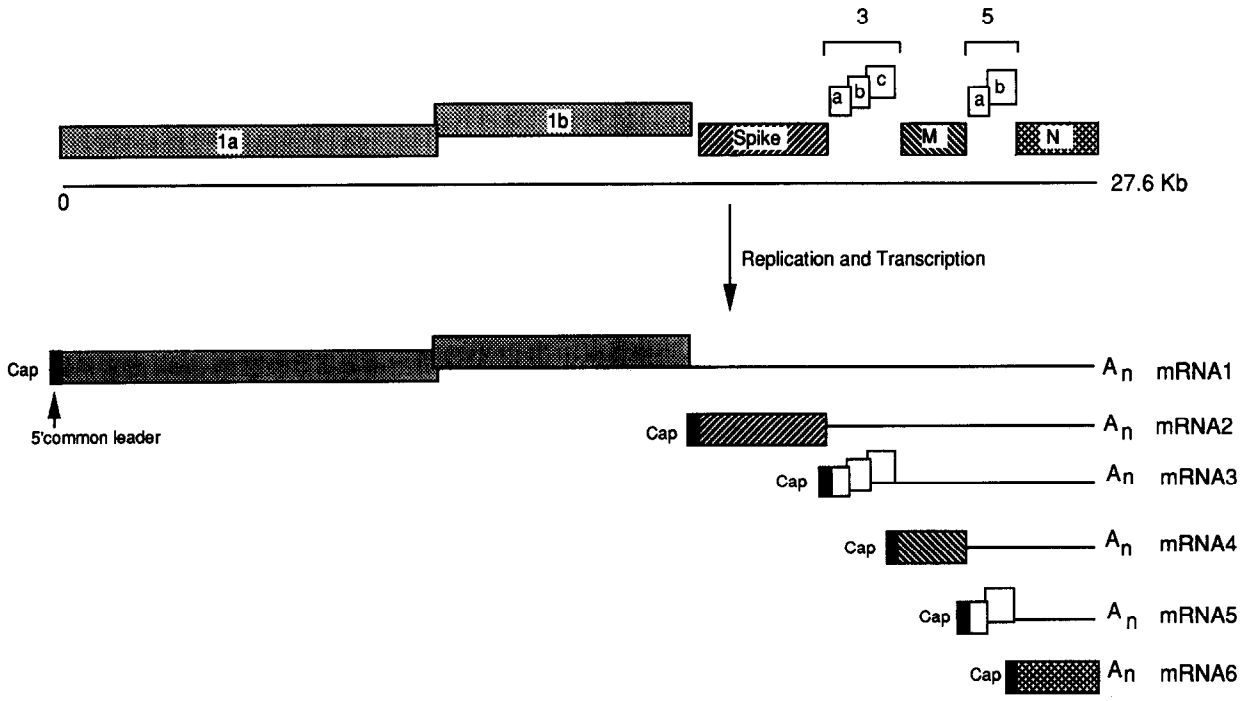


FIG. 1. Organization and expression of the IBV genome and its subgenomic mRNAs. Major open reading frames are boxed.

and Gough, 1971); UK/68/84 and Portugal/322/82, both provided by Dr. J. K. A. Cook, AFRC Institute of Animal Health, Houghton Laboratory (Houghton, Cambridgeshire, UK).

Nucleotide sequencing of genomic RNA

Virus RNA extraction was based on the guanidinium thiocyanate method of Chomczynski and Sacchi (1987). An extraction solution was prepared by mixing at 65°: 250 g guanidinium thiocyanate (Fluka); 17.6 ml of 0.75 *M* sodium citrate, pH 7; 26.4 ml of 10% sarcosyl; 293 ml of water. On the day of use 72 μl of 2-mercaptoethanol was added per 10 ml of the extraction solution. Clarified allantoic fluid from eggs infected with IBV was centrifuged at 34,000 *g* for 90 min to pellet virions. The pellet was resuspended in 2 ml of extraction solution and the following solutions were added sequentially: 0.2 ml 2 *M* sodium acetate, pH 4.0; 2 ml phenol; 0.4 ml chloroform/isoamyl alcohol mixture (49:1). The mixture was shaken for 5 min and centrifuged at 10,000 *g* for 20 min at 4°. RNA was recovered from the aqueous phase by ethanol precipitation and centrifugation. After addition of 0.6 ml of extraction solution, the RNA was reprecipitated with ethanol, recovered by centrifugation, washed, dried, and dissolved in diethylpyrocarbonate-treated water.

The RNA was sequenced using oligonucleotide primers as previously described (Cavanagh and Davis, 1988).

Analysis and purification of bacterial fusion proteins

Analysis and purification of bacterial fusion proteins were carried out as described by Smith *et al.* (1990). Briefly, bacteria carrying pEX-based expression plasmids were grown at 30° and then induced to produce the fusion protein by raising the temperature to 42°. Following induction, the bacteria were harvested, treated with lysozyme, and lysed in buffer containing Triton X-100. Insoluble material containing the bulk of the fusion protein was recovered by centrifugation and applied to a preparative 7.5% SDS–polyacrylamide gel. Bands corresponding to the fusion protein were located by staining in water with Coomassie brilliant blue and excised, and the fusion protein was recovered by electroelution.

Preparation of antifusion protein antibodies

Rabbits were injected intramuscularly with about 150 μg of purified fusion protein emulsified in Freund's complete adjuvant and boosted at monthly intervals by the same route with similar doses of protein emulsified in Freund's incomplete adjuvant. Three weeks after each boost, blood was collected from the ear vein, and

the specificity and reactivity of the antiserum were tested by immunoprecipitation against radiolabeled antigens synthesized by *in vitro* translation of *in vitro* synthesized mRNAs.

Labeling of IBV-infected cells with [^{35}S]methionine and [^{35}S]cysteine

Confluent monolayers of CK cells were infected with IBV at a multiplicity of infection (m.o.i.) of approximately 1 plaque-forming unit (PFU)/cell. Before labeling, the cells were incubated in medium free of methionine and cysteine for 30 min. After 120 min labeling with [^{35}S]-methionine (100 μCi/ml) and [^{35}S]cysteine (100 μCi/ml), the cells were scraped off the dishes in saline, recovered by centrifugation, and stored at −70° until required.

Radioimmunoprecipitation

IBV-infected CK cells were lysed with RIPA buffer (50 m*M* Tris–HCl, pH 7.5, 150 m*M* NaCl, 1% sodium deoxycholate, 0.1% SDS) and precleared by centrifugation at 200,000 *g* for 30 min at 4°. Immunoprecipitation was carried out as described previously (Brierley *et al.*, 1987).

Transient expression of mRNA3 products in eukaryotic cells

Open reading frames placed under control of the T7 promoter were expressed transiently in eukaryotic cells as described by Fuerst *et al.* (1986). Briefly, semiconfluent monolayers of CV-1 cells were infected with 25 PFU/cell of a recombinant vaccinia virus (vTF7-3) which expresses the T7 phage RNA polymerase and then transfected with appropriate plasmid DNA by the standard calcium phosphate-mediated procedure. After 30 min at room temperature, fresh medium was added and the cells were incubated at 37° for 2 hr. The medium was removed and replaced with methionine-free medium supplemented with 1% fetal calf serum and 25 μCi/ml [^{35}S]methionine. The radiolabeled cells were harvested at 18 hr postinfection.

SDS–polyacrylamide gel electrophoresis

SDS–polyacrylamide gel electrophoresis (PAGE) of virus polypeptides was carried out using 20 or 22% polyacrylamide gels (Laemmli, 1970) with an acrylamide:bis acrylamide ratio of 300:1 (20% gel) and 440:1 (22% gel) according to previously described procedures (Inglis *et al.*, 1976). Labeled polypeptides were detected by autoradiography or fluorography of dried gels.

Cell-free transcription and translation

In vitro transcription from plasmid DNA using SP6 or T7 phage RNA polymerase was carried out as described by Brierley *et al.* (1987), incorporating the dinucleotide ^{7m}GpppG to provide a 5′ cap structure (Contreras *et al.*, 1982). Product mRNA was recovered from the reactions by extraction with phenol–chloroform (1:1) and precipitation with ethanol. The RNA was further purified by gel filtration on Sephadex G50 and translated in the wheat germ cell-free system in the presence of 0.75 μCi/μl [^{35}S]methionine as described previously (Inglis *et al.*, 1977). Reaction products were separated by polyacrylamide gel electrophoresis and detected by autoradiography.

Construction of plasmids

The m13 clone 141 (a kind gift from Dr. M. Boursnell), which includes the 3a, 3b, and 3c regions as part of a larger IBV genomic sequence (containing IBV genomic RNA sequence from nucleotide 23,358 to 24,706), was altered by *in vitro* mutagenesis to introduce a cleavage site for the restriction enzyme *Bgl*II 20 nucleotides upstream of the 3a initiation codon (Fig. 2). A cDNA fragment (910 bp) containing the 3a–b–c region together with the beginning of the IBV M gene was then excised with *Bgl*II and *Hin*dIII (which cuts within the m13 vector sequence immediately adjacent to the IBV sequence 173 nucleotides (excluding the GC tail sequence) downstream of the 3c ORF) and cloned into *Bgl*II- and *Hin*dIII-digested pING14 to give the plasmid pING16 (Fig. 2). Plasmid pING14 consists of a modified form of pSP64T (Krieg and Melton, 1984), which contains the intergenic region of filamentous bacteriophage f1 (Dotto *et al.*, 1981) inserted at a unique *Pvu*II site (Inglis, unpublished data). Thus pING16 can be converted to single-stranded DNA by superinfection of plasmid-bearing cells with the bacteriophage R408 (Russel *et al.*, 1986).

Plasmid pIBT1 contains only the 3c ORF and was made by cloning a *Pvu*II–*Hin*dIII restriction fragment (563 bp) from M13 clone 141 into *Sma*I–*Hin*dIII-digested plasmid pT7-1 (Genescribe). Plasmid pIBT2 contains only the 3b and 3c ORFs and was constructed by cloning a 726-bp *Xba*I–*Hin*dIII restriction fragment from m13 clone 141 into *Xba*I–*Hin*dIII-digested pT7-1. Plasmid pIBT3 contains each of the 3a, 3b, and 3c ORFs and was made by cloning a *Bgl*II–*Hin*dIII restriction fragment from pING16 into *Bam*HI–*Hin*dIII-digested pT7-1. The correct structure of these three plasmids was confirmed by nucleotide sequencing.

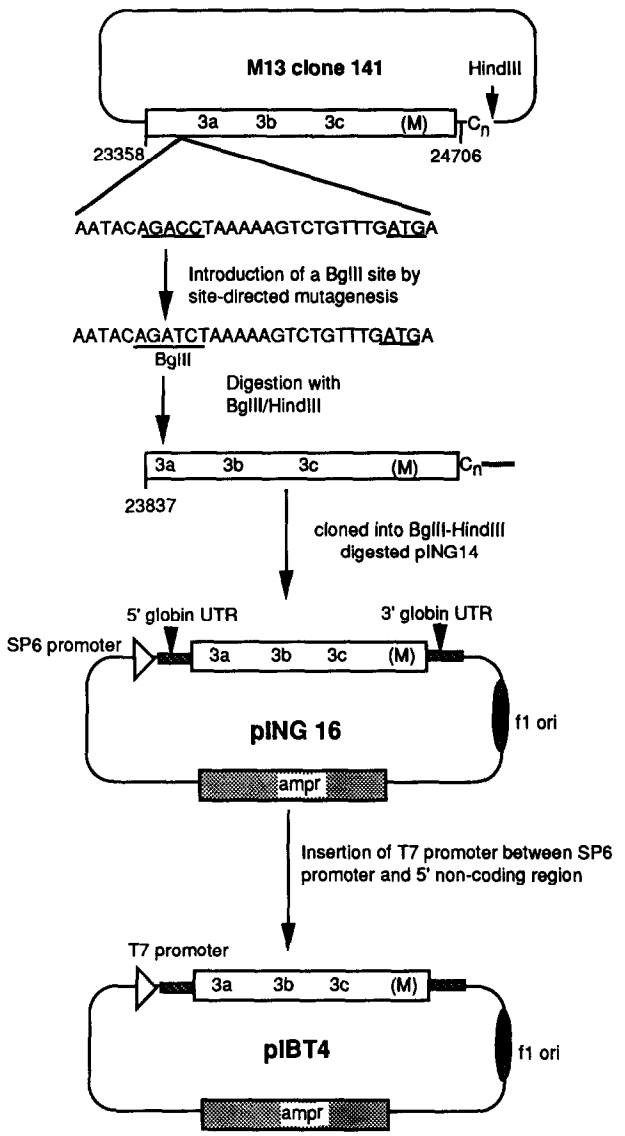


FIG. 2. Construction of plasmids pING16 and pIBT4 (see Methods).

Plasmid pIBT4 was constructed by inserting, using site-directed mutagenesis, a sequence corresponding to the T7 RNA polymerase promoter (AATTAATACGACTCACTATAGGGAGA) just downstream of the SP6 promoter and 105 nucleotides upstream of the initiation codon of 3a in pING16 (Fig. 2). This plasmid was further modified to produce three new plasmids (pIBM6, pIBT12, and pIBM5).

Plasmid pIBM6, in which the 3a, 3b, and 3c ORFs are fused to form a single continuous ORF, was created in two stages by site-directed mutagenesis. In the first, a single additional A residue was inserted 27 nucleotides upstream of the 3b initiation codon using the oligonucleotide 5'-TTACAGTCTAGAACTGACC-3'. In the second, an additional two nucleotides (GA) were inserted 23 nucleotides upstream of the initiation codon for 3c using the oligonucleotide 5'-TTGGGAAATAGAGAGTCAGC-3'.

Plasmid pIBT12 which contains the authentic IBV leader sequence upstream of the 3a–b–c coding region (replacing the 5' noncoding region of β-globin derived from pSP64T) was also constructed from pIBT4 by site-directed mutagenesis in two stages. The 5' globin noncoding sequence of pIBT4 was first replaced with the first half of the IBV leader sequence, and the second half of the IBV leader was then inserted to give the plasmid pIBT12. A derivative of this plasmid, pIBT13, was constructed by excising from pIBT12 a *Bgl*II–*Hin*dIII fragment containing the 3a–b–c coding region and replacing it with a *Bgl*II–*Hin*dIII fragment derived from the plasmid pIBM6, which contained the fused 3abc ORF.

Plasmid pIBM5 was prepared from pIBT4 by altering, through site-directed mutagenesis, a potential stop codon (TAA), which lies 31–35 nucleotides upstream of the 3a initiation codon, to the codon TCA. This plasmid was digested with *Bgl*II and *Pvu*II to generate a 349-bp fragment comprising the 3a ORF together with most of 3b. This fragment was then cloned into pEX2 (Stanley and Luzio, 1984), which had been first digested with *Pst*I, end-repaired with DNA *Pol*I, and then redigested with *Bam*HI, to give the plasmid pEX2/3a. A 685-bp fragment containing the complete 3b ORF was excised from pIBT4 by digestion with *Xba*I, end-repair using the T4 DNA polymerase, and redigestion with *Pst*I before cloning into *Sma*I- and *Pst*I-digested pEX1 (Stanley and Luzio, 1984) to generate the plasmid pEX1/3b (Fig. 4a).

Site-directed mutagenesis

Site-directed mutagenesis of single-stranded plasmid DNA was carried out essentially as described by Brierley *et al.* (1989).

RESULTS

Nucleotide sequencing of the 3a–b–c region from different IBV strains

If each of the 3a, 3b, and 3c ORFs identified in the genome of the Beaudette strain of IBV represents a genuine virus gene, it would be expected that each would be conserved in other strains of the virus. For this reason we deduced, by direct RNA sequencing using synthetic oligonucleotide primers, the nucleotide sequence of the equivalent genomic region from an additional four virus strains, M41, 322/82, 68/84, and

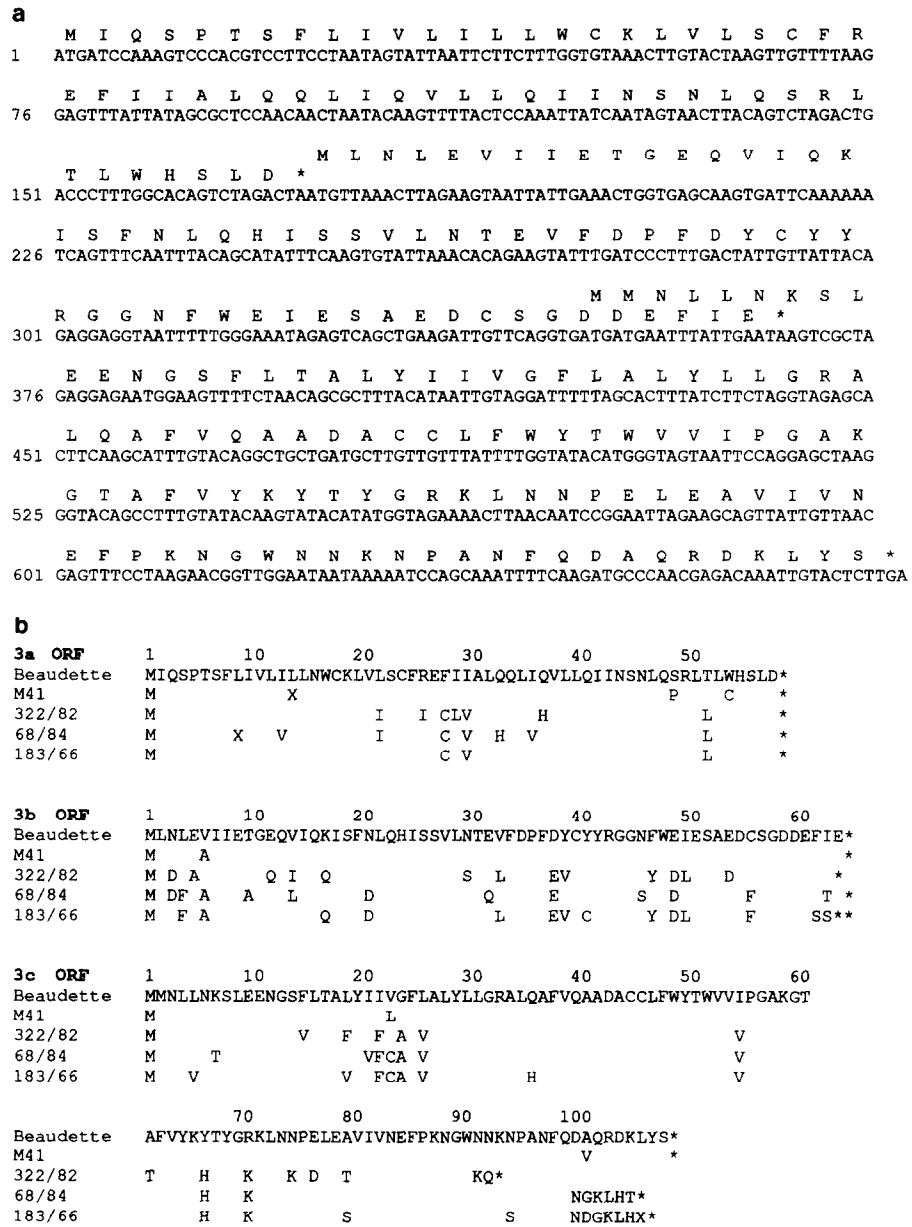


FIG. 3. (a) Nucleotide sequences and the predicted amino acid sequences of the 3a, b, and c ORFs from IBV strain Beaudette (Boursnell *et al.*, 1985). (b) Predicted amino acid sequences of the 3a, 3b, and 3c ORFs from five different strains of IBV: Beaudette; M41/41; UK/68/84; UK/183/66; and Portugal/322/82. Only amino acids found to differ from the Beaudette sequence are shown. 'X' means that the nucleotide sequence was not sufficiently clear to identify unambiguously the amino acid codon. Translational stop codons are indicated by the symbol *.

183/66. The results indicated that all three ORFs are conserved in each strain, suggesting that they do indeed encode virus proteins. The predicted amino acid sequence of these proteins, compared with the Beaudette strain (Boursnell *et al.*, 1985), is shown in Figs. 3a and 3b, and their comparative features are listed in Table 1. It is evident that the 3a ORF is highly conserved in position, length, and sequence similarity. The 3a ORFs of the Beaudette and M41 strains are particularly closely related, with the other strains appearing to form a rather more distant group. The amino acid variation observed in these strains, however, did not in general alter the overall hydrophobic character of the protein observed originally by Boursnell *et al.* (1985). The 3b ORF, although also similar in position and length in each case, showed greater diversity in sequence, with the 322/82, 68/84, and 183/66 strains again appearing closely related to each other, but up to 22% different

TABLE 1

COMPARISON OF THE POSITION OF TRANSLATION STOP CODON AND THE NUMBERS (%) OF AMINO ACID DIFFERENCES IN mRNA3 OF FIVE DIFFERENT STRAINS OF IBV

	Position of translation stop codon in ORF[a]			Number (%) amino acid differences in ORF[a]		
IBV strain	3a	3b	3c	3a	3b	3c
Beaudette	59	65	110	—	—	—
M41	59	65	110	2 (3)	1 (2)	2 (2)
322/82	59	65	94	6 (10)	13 (20)	14 (14)
68/84	59	64	107	7 (12)	12 (19)	13 (12)
183/66	59	64	108	4 (7)	14 (22)	16 (15)

[a] Relative to the IBV strain Beaudette ORFs. All the ORFs start at the same position as in IBV strain Beaudette.

from the Beaudette sequence which proved very similar to that from M41. A similar picture emerged from comparative analysis of the 3c ORFs, with the Beaudette and M41 strains once again forming rather distinct group from the others. In this case, however, although the beginning of the ORF (i.e., the first Met codon) occupies an identical genetic location in each case, the point at which the ORF terminates is somewhat variable, with the 322/82 ORF truncated by 16 amino acids in comparison to the Beaudette strain, and the 68/84 and 183/66 ORFs missing 3 and 2 amino acids, respectively. Since a 3c-encoded protein has already been identified in cells infected with the Beaudette strain of IBV (Smith *et al.*, 1990), these data suggest that the extreme C-terminus of the protein is unnecessary for its function.

Production of monospecific antibodies against the predicted products of the 3a and 3b ORFs

We next set out to identify the protein products of the 3a and 3b ORFs through expression of their sequences in bacteria and preparation of monospecific antisera against the resulting protein products. The bacterial expression system chosen (based on the PEX series of plasmids) was that developed by Stanley and Luzio (1984); foreign sequences are inserted in the correct reading frame, at the C-terminus of a β-galactosidase gene, which is itself fused to the promoter, operator, and N-terminal region of the cro gene of bacteriophage λ. Synthesis of the β-galactosidase fusion protein in bacteria carrying these plasmids is therefore repressed in cells carrying a functional λ repressor, but can be induced in *Escherichia coli* POP2136 cells (which carry a temperature-sensitive repressor) simply by increasing the culture temperature to 42°. We previously used an identical strategy to identify the protein product of the 3c ORF (Smith *et al.*, 1990).

A cDNA fragment containing the entire 3a ORF was cloned into plasmid pEX2, giving the plasmid pEX2/3a (Methods and Fig. 4a) such that the β-galactosidase and 3a coding sequences were contiguous. Similarly, a fragment containing the 3b ORF was cloned into pEXI, giving plasmid pEX1/3b. (Methods and Fig. 4a). The continuity of reading frames between the bacterial and the viral sequences was checked by nucleotide sequencing. Following heat induction, the proteins produced by host bacteria carrying these plasmids were examined by gel electrophoresis (Fig. 4b). In each case, a fusion protein of the expected size was produced in place of the wild-type unfused protein. These fusion proteins were purified by preparative gel electrophoresis and inoculated into rabbits to produce specific antisera (Methods).

The specificity of the polyclonal antisera generated by this route was then tested by immunoprecipitation studies using radiolabeled 3a- and 3b-encoded proteins synthesized by *in vitro* translation of artificial mRNA (see later sections). The results of this experiment (data not shown) indicated that the sera were indeed capable of recognizing specifically the appropriate target sequence.

Identification of gene products encoded by the 3a and 3b ORFs in IBV-infected cells

The antisera raised against the 3a and 3b specific bacterial fusion protein were next tested in immunoprecipitation experiments against [^{35}S]methionine- and [^{35}S]cysteine-labeled lysates prepared from IBV-infected CK cells (Fig. 5). Although a relatively high background of nonspecific precipitation was observed, the anti-3a serum detected a weak band with the electrophoretic migration expected for 3a, as judged by comparison with *in vitro*-translated 3a. That this does indeed represent the product of the 3a ORF is suggested by the fact that this protein was not precipitated from the same lysate by either preimmune serum or by hyperimmune sera directed against other PEX fusion protein. In addition, it was not detected in mock-infected control lysate using either specific or nonspecific antisera. The protein was not detectable in unprecipitated infected cell lysates, even after a long exposure, suggesting that it is present in very small amounts. The anti-3b antiserum was likewise able to precipitate, although very weakly, a small protein of the size ex-

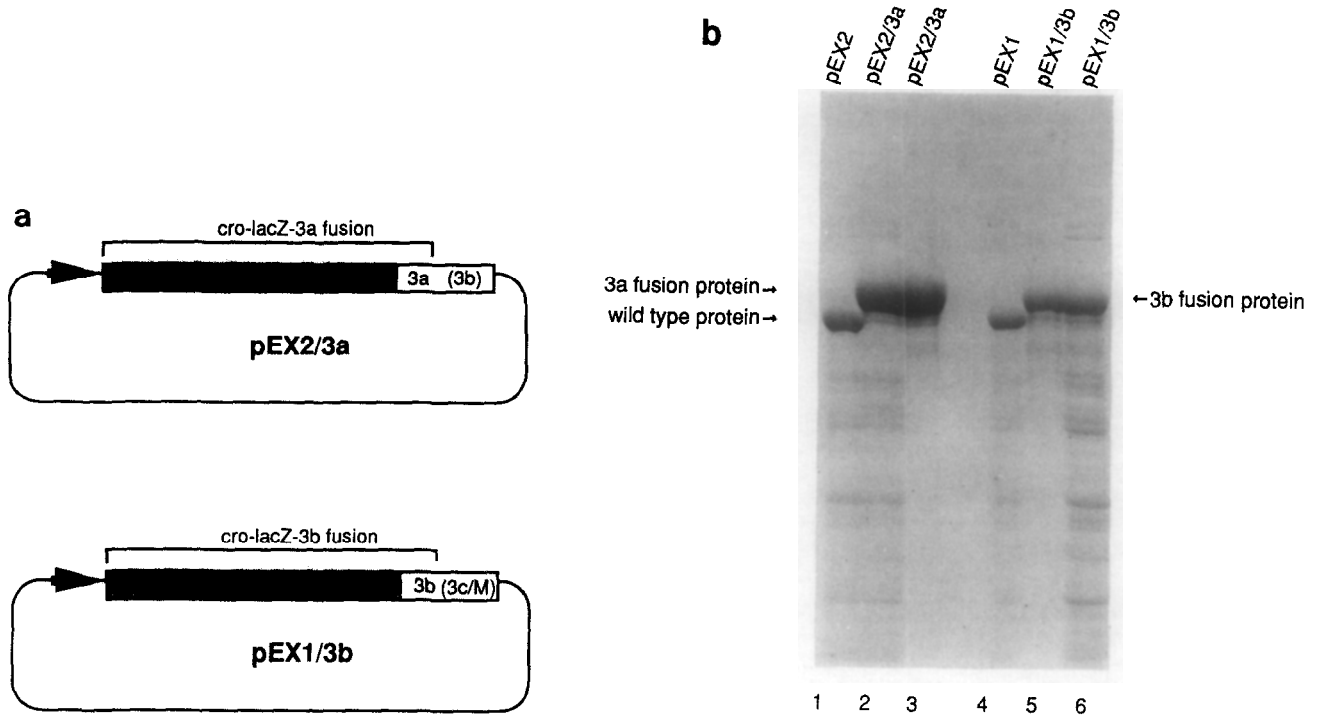


FIG. 4. (a) Diagram of the plasmids pEX2/3a and pEX1/3b constructed to express the 3a or 3b ORF in *E. coli* as a fusion protein with β-galactosidase (see Methods). (b) Proteins induced in *E. coli* (POP2136) cells carrying plasmid pEX1, pEX2, pEX2/3a, or pEX1/3b after incubation at 42°. Crude cell lysates (lanes 1, 2, 4, and 6) or purified polypeptides (lanes 3 and 5) were separated on a 7.5% SDS–polyacrylamide gel and detected by staining with Coomassie brilliant blue.

pected for the 3b product as judged by its coelectrophoresis with *in vitro*-translated 3b. Once again comparison with control immunoprecipitations from infected and mock-infected cells supports the idea this is indeed the 3b product. Thus it appears that both the 3a and the 3b ORFs are indeed expressed in IBV-infected cells.

Polypeptides 3a, 3b, and 3c can be produced from a single mRNA

Expression of each of the 3a, 3b, and 3c ORFs in IBV-infected cells would imply that all three are translated from mRNA3, since none are present on the next smallest member of the nested set of subgenomic mRNAs. To investigate this possibility directly, mRNAs containing the three ORFs were synthesized artificially using the T7 phage RNA polymerase and tested in the wheat germ cell-free translation system for their ability to direct synthesis of each protein. Plasmids pIBT4, pIBT5, and PIBT6 (Fig. 6a) were digested with *Hin*dIII and transcribed *in vitro* using T7 RNA polymerase. The resulting transcripts were then translated in wheat germ cell-free extracts (Fig. 6b). RNA transcripts from *Hin*dIII-digested pIBT4 plasmid DNA, which should contain each of the 3a, b, and c ORFs, directed the synthesis of a protein of MW approximately 12K, corresponding to the 3c product (Smith *et al.*, 1990) and in addition a large amount of a smaller proteins of approximately 7K, the expected MW of the 3a products. A faint band migrating slightly more slowly than this protein was also observed; this protein is of the expected size for a product of the 3b ORF. RNA transcripts from *Hin*dIII-digested pIBT5, which contained only the 3b and 3c ORFs, directed the synthesis of two products migrating at the positions of 3b and 3c as above. Similarly, one major product, which migrated at the position of 3c, was produced from pIBT6 transcripts. Further evidence for the identity of these products was obtained by translation of mRNA from *Hin*dIII-, *Pvu*II-, and *Xba*I-digested pIBT4. As predicted, the 12K protein was not synthesized in response to transcripts from *Pvu*II-digested pIBT4 DNA (containing only the 3a and b ORFs), suggesting that it does indeed represent the products of 3c (Fig. 6b). Translation of RNA from *Xba*I-digested pIBT4DNA (containing the 3a ORF alone) produced only a small protein of approximately 7K, confirming its identity as a product of the 3a ORF. During the course of these experiments, we constructed an additional series of plasmids, bearing 3c alone, 3b

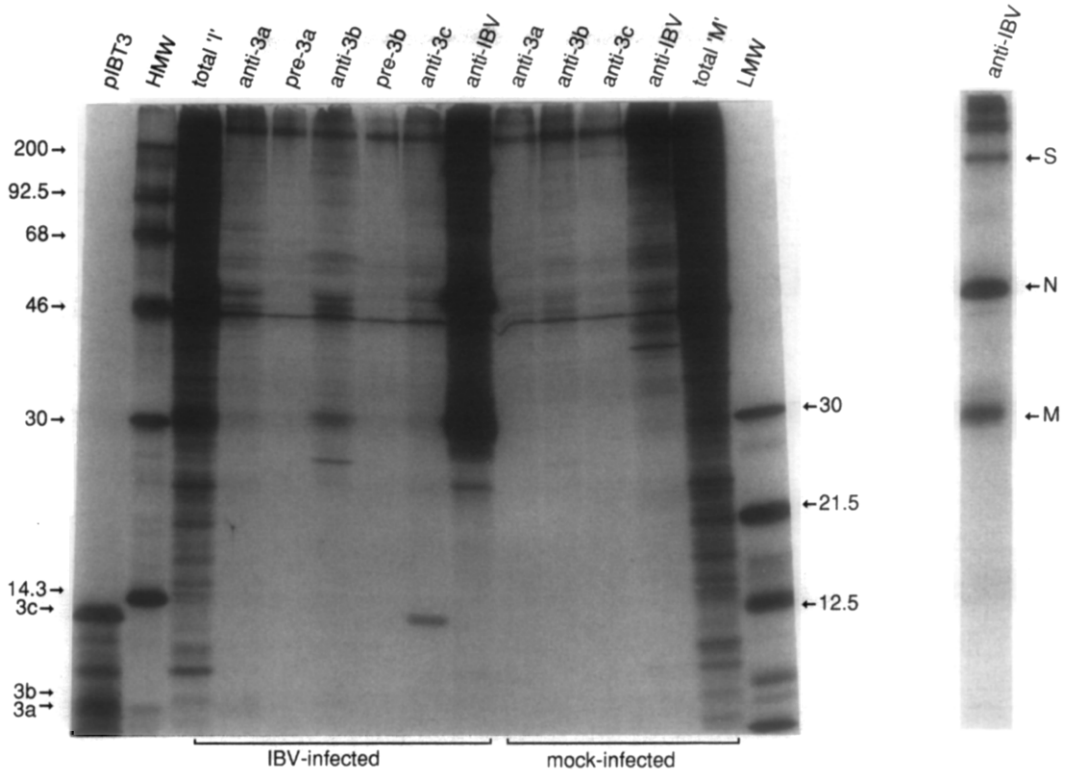


FIG. 5. Detection of polypeptides encoded by the 3a, b, and c ORFs in IBV-infected and mock-infected CK cells by immunoprecipitation using antisera raised against bacterial fusion proteins. Cells were labeled with [^{35}S]methionine and [^{35}S]cysteine, lysates prepared, and polypeptides either analyzed directly or immunoprecipitated with the antisera indicated above each lane. Lanes labeled LMW and HMW represent low- and high-molecular-weight markers, respectively. Polypeptides were separated on a 22% SDS–polyacrylamide gel and detected by fluorography.

+ 3c, and 3a + 3b + 3c (pIBT1, pIBT2, and pIBT3, respectively, Methods), which differed in that they were based on a different T7 promoter-containing plasmid and lacked any globin sequences upstream or downstream of the IBV ORFs. Translation of mRNAs transcribed from these plasmids is shown in Fig. 6c. Once again the polypeptides produced were of the expected sizes for the products of the 3a, 3b, and 3c ORFs. The translation product of pIBT1-derived RNA included two minor species running just below 3c; the higher of these bands appears to represent a product derived by internal initiation from the residual portion of the M gene present at the 3′ end of the transcript, since it was not seen when shorter transcripts were translated; the lower of the two probably represents a premature termination product related to 3c. One surprising aspect of these results was that the relative proportions of the proteins synthesized in response to pIBT3-derived mRNA differed markedly from those observed on translation of mRNA from pIBT4 in that with pIBT3, the 3c protein was clearly the most abundant product, initiation at the 3a and 3b ORFs occurring very inefficiently. Other studies underway in this laboratory suggest that this difference may be due to the fortuitous formation of some kind of secondary or tertiary RNA structure by the vector-derived 5′ noncoding sequence of this particular mRNA, which blocks synthesis of 3a and 3b, but not 3c; the effect was not seen with synthetic tricistronic mRNAs containing other 5′ noncoding sequences, and it could be abolished by a number of single nucleotide substitutions near the 5′ end (data not shown). Further work is in progress to investigate this phenomenon.

Interpretation of the translational patterns produced by these tricistronic mRNAs, however, is complicated by the close similarity in molecular weight between the 3a- and the 3b-encoded products and the difficulty in routinely separating the two proteins by gel electrophoresis. In order to facilitate further studies on initiation at each of the three ORFs, we therefore altered pIBT4 by site-directed mutagenesis to fuse the 3a, 3b, and 3c ORFs together (pIBM6), thereby creating one continuous 3abc ORF. Since the ORFs overlap to some extent, in different reading frames, this can be accomplished through the introduction of a single additional nucleotide within the overlap between 3a and 3b and two

additional nucleotides within the overlap between 3b and 3c, at positions which would be unlikely to influence the efficiency of initiation at the subsequent start codon (see Methods and Fig. 6a). The rationale for this is that if translation of the 3a, 3b, and 3c ORFs can be initiated separately as expected, a mRNA bearing this continuous ORF should yield three new products with MW 27K (corresponding to the 3abc fusion protein), 20K (3bc fusion protein), and 12.4K (3c protein alone). Figure 6b shows that mRNA transcribed from *Hin*dIII-digested pIBM6 does indeed encode polypeptides of the expected MW. Moreover, the proportions of the three products synthesized in response to this mRNA, taking into account the number of methionine residues carried by each, are broadly similar to those of the separate proteins translated from pIBT4-directed mRNA.

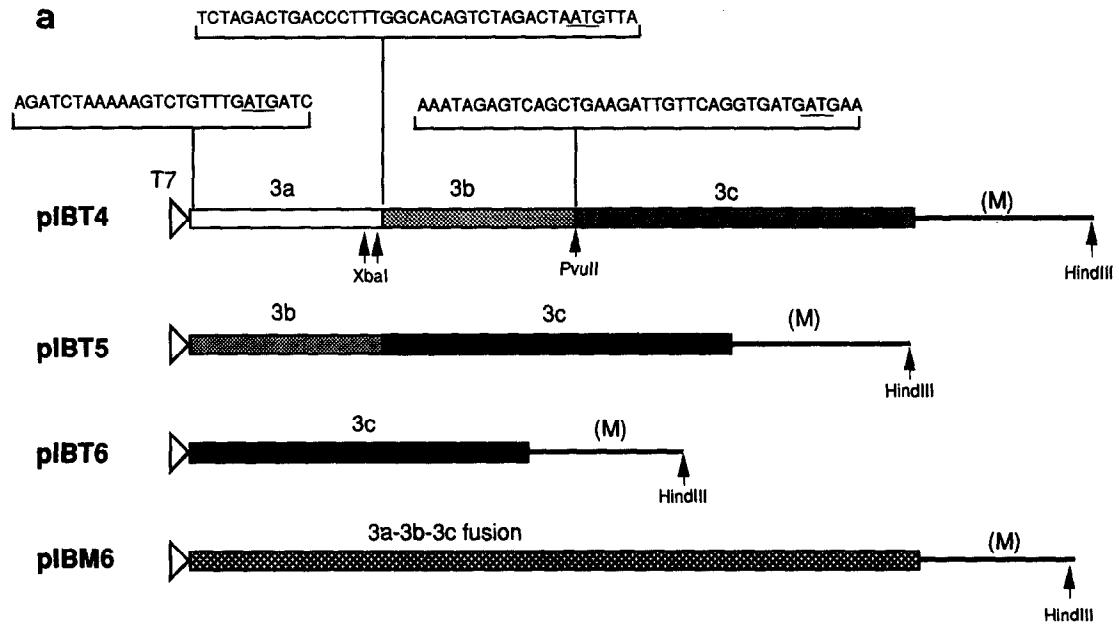
These results therefore demonstrate clearly that a single mRNA bearing the 3a, b, and c ORFs can direct the synthesis of all three polypeptides *in vitro,* implying that the natural virus mRNA3 could be functionally tricistronic. However, natural IBV virus mRNAs differ significantly from the synthetic transcripts studied in these experiments in that they contain at their 5′ end a virus-encoded leader sequence of untranslated nucleotides (Brown *et al.,* 1984), whereas the synthetic mRNAs described so far contain a 5′ untranslated region (UTR) derived from the rabbit β-globin gene (provided by the cloning vector). Such a difference could conceivably affect translational initiation.

Effect of IBV leader sequence on translation of the 3 a, b, and c ORFs

The IBV mRNAs contain a common 5′ leader sequence of about 60 nucleotides (Brown *et al.,* 1984). Available evidence suggests that the leader RNA present on each subgenomic mRNA is added by a unique mechanism of "leader primed transcription," in which leader RNA is transcribed from a (−) stranded RNA template, dissociates from the template, and then rebinds at certain sites downstream to serve as the primer for transcription of subgenomic mRNAs (Lai, 1986). The point at which the leader joins the body of the subgenomic mRNA is characterized by a core region of homology (CTTAACAA), and for mRNA3 the precise sequence at the junction between the leader and the body has been established (Brown *et al.,* 1986). We therefore wished to investigate the effect on the translational capacity of the artificial IBV mRNA3 of an authentic IBV 5′ leader sequence. This seemed particularly important in view of the effect of the 5′ UTR observed previously on the relative translational efficiency of the 3a, 3b, and 3c ORFs (cf. Figs. 6b and 6c). We therefore introduced the authentic 5′ leader sequence into plasmid pIBT4 by site-directed mutagenesis (pIBT12—see Methods and Fig. 7a) and translated a tricistronic mRNA transcribed from this plasmid in the wheat germ system as before (Fig. 7b). The results show that the 3a-, 3b-, and 3c-encoded polypeptides were synthesized in proportions similar to those produced by pIBT4-derived transcripts. An equivalent plasmid, in which the 3a, b, and c ORFs had been fused into one continuous 3abc ORF (pIBT13) using the same procedure as described above for construction of pIBM6 from pIBT4, was also used to prepare tricistronic mRNA for *in vitro* translation. Once again the translational pattern was similar to that produced by the equivalent RNA lacking the leader sequence (transcribed from pIBM6). Thus the presence of the IBV leader sequence does not appear to influence the capacity of the mRNA to direct synthesis of three separate polypeptides.

Expression of mRNA3-encoded polypeptides *in vivo*

We finally tested the ability of a synthetic transcript carrying the 3a, b, and c ORFs to function as tricistronic mRNAs in intact eukaryotic cells. In order to carry out this experiment, we used the system described by Fuerst *et al.* (1986), in which CV1 cells are infected with a vaccinia recombinant virus expressing the T7 phage RNA polymerase and subsequently transfected with plasmid DNA containing a test gene under the control of the T7 RNA polymerase promoter. We therefore tested directly for expression in intact cells of the 3a, b, and c ORFs using the T7 promoter-based plasmids described above. For these experiments we used only the plasmid which contained the single fused 3abc ORF together with the authentic IBV leader sequence (pIBT13). Thus translation products initiating from each ORF could be detected simultaneously by immunoprecipitation using antiserum directed against the 3c polypeptide (Smith *et al.,* 1990), eliminating the need for the use of different antisera with unavoidable differences in avidity. The result of this experiment is shown in Fig. 8. Transient expression of a protein with the characteristics of 3abc was clearly detectable in cells transfected with pIBT13, but not in untransfected cells. In addition a protein corresponding to the 3bc ORF was also produced, indicating that initation of translation had occurred at both the 3a and the 3b start sites. It was more difficult to detect a product initiating at the 3c start site, although a faint band was observed (migrating just below the background band). However, since 3c is definitely expressed in infected cells (Fig. 5 and Smith *et al.,* 1990),



Expected translation products			
pIBT4/HindIII	6.8K (3a)	7.4K (3b)	12.4K (3c)
pIBT4/PvuII	6.8K (3a)	6.2K (3b')	
pIBT4/XbaI	6.0K (3a')		
pIBT5/HindIII		7.4K (3b)	12.4K (3c)
pIBT6/HindIII			12.4K (3c)
pIBM6/HindIII	26.7K (3abc)	19.9K (3bc)	12.4K (3c)

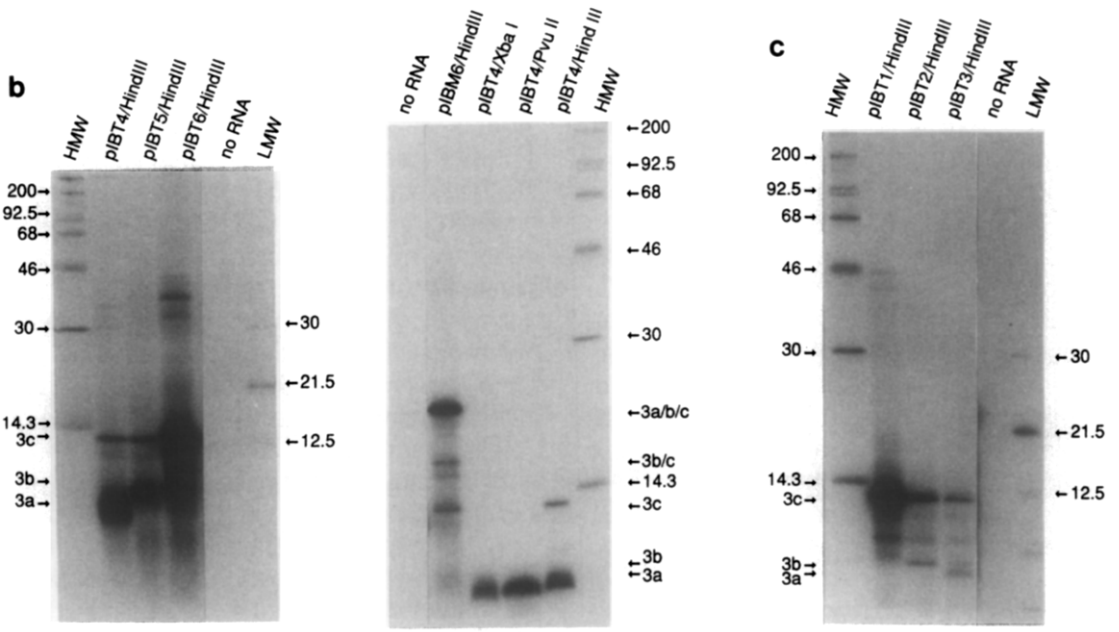


FIG. 6. (a) Diagram of plasmids pIBT4, pIBT5, pIBT6, and pIBM6 showing the T7 promoter site, the sequence context around the initiation codon for the 3a, b, and c ORFs, and the restriction enzyme sites used to linearize the plasmids for *in vitro* transcription. Also shown are the sizes of the *in vitro* translation products expected from initiation at the beginning of each of the three ORFs on mRNA transcribed from plasmids

it seems clear that the 3c start site can function on a polycistronic mRNA in intact cells. It therefore appears that a mRNA carrying all three mRNA3 ORFs can function polycistronically *in vivo*.

DISCUSSION

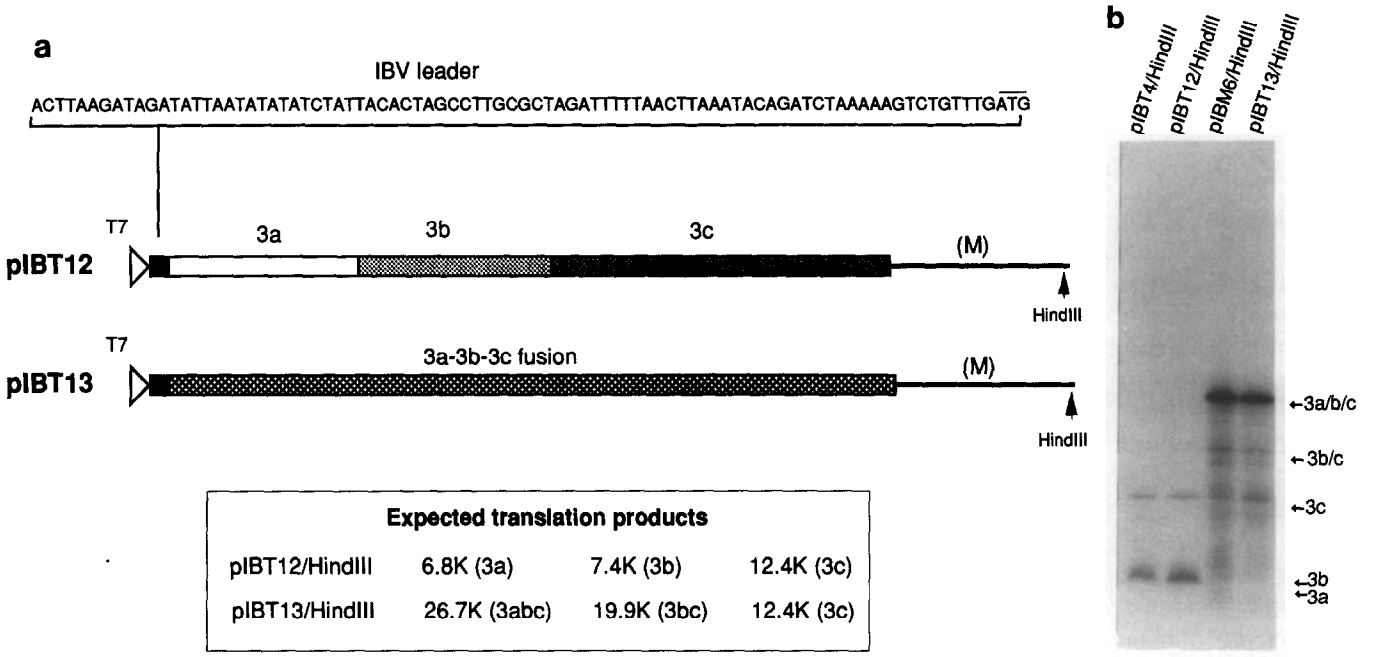
Nucleotide sequence analysis of the Beaudette strain of IBV, carried out by Boursnell *et al.* (1985), indicated that the region of the RNA genome corresponding to the 5′ "unique" portion of IBV mRNA3 contained three small ORFs, with the potential to encode previously unidentified virus proteins. We showed previously that the most distal of these three ORFs, 3c, does indeed encode a polypeptide of 12.4K (Smith *et al.*, 1990), and the results reported here indicate clearly that the 3a and 3b ORFs also represent genuine virus genes. This conclusion is based on nucleotide sequence analysis of four additional strains of IBV, which showed that closely related ORFs with similar size and location are conserved in each case, and on direct identification of 3a- and 3b-encoded proteins in IBV-infected cells using monospecific antisera. These novel proteins are of the expected size for the products of the 3a and 3b ORFs, that is 6.7K and 7.4K, respectively, and furthermore they comigrate during SDS–gel electrophoresis with the 3a and 3b proteins synthesized by *in vitro* translation of synthetic mRNAs. Thus all three ORFs are expressed in infected cells.

Available evidence from direct studies of IBV-infected cell mRNAs implies that each of the 3a, 3b, and 3c polypeptides must be expressed from IBV mRNA3, a 3.8-kb subgenomic mRNA consisting of a 65-nucleotide 5′ leader sequence (corresponding to the 5′ end of the genomic RNA) fused to a "body" which initiates 23832 nucleotides from the genomic 5′ end and runs through to its extreme 3′ terminus. This is based on the general observation that coronavirus mRNAs form a 3′ coterminal nested set in which translation is restricted to the 5′ regions that are not represented on the next smallest mRNA species (Stern and Sefton, 1984); after mRNA3, the next smallest subgenomic mRNA, mRNA4, which encodes the virus membrane protein, does not contain any of the 3a, b, and c ORFs.

The implication that mRNA3 can encode three different virus polypeptides is considerably strengthened by our finding that synthetic RNA transcripts containing all three ORFs, either as separate cistrons or as part of a single continuous ORF, can direct the synthesis in wheat germ cell-free extracts, and in intact cells, of products from all three reading frames. This was true whether or not the artificial mRNAs contained the authentic IBV leader sequence at their 5′ end. It seems highly likely therefore that mRNA3 is functionally tricistronic.

How might the expression of all three ORFs be achieved from a single species of mRNA? The pattern of translation observed *in vitro* from mRNAs containing all three ORFs suggested that initiation occurred most efficiently on the ORF nearest the 5′ end of the mRNA, 3a, with the two downstream ORFs being translated much less frequently. The sequence around the 3a start codon (CUGUUUGAUGAUC) has U at position −3 and A at +4; around the 3b start codon (UAGACUAAUGUUA) it has C at position −3 and U at +4; and around the 3c start codon (UCAGGUGAUGAUG) it has a G at position −3 (for both possible start codons) and A at +4. According to Kozak's scanning model for translational initiation (Kozak, 1989), the utilization of an AUG codon as an initiation signal for translation is determined both by its location within the mRNA molecule and by the sequences surrounding it (sequence context). From a survey of known initiation signals, Kozak (1987) concluded that the optimal sequence context for initiation of translation is 5′-CCACCAUGG, with the presence of a purine residue at the −3 and +4 positions being of primary importance for the efficient recognition of the AUG as initiation signal. However, if the sequence context around the AUG is suboptimal, ribosomes may fail to initiate and could therefore scan past, initiating instead at AUG codons located downstream. By these criteria, the 3a and 3b initiation codons are in a relatively poor context for ribosome recognition, while the 3c initiation codon is in a relatively good context. It seems therefore that a "leaky scanning" model (Kozak, 1989) could explain the pattern of expression of mRNA3; most ribosomes scanning from the 5′ end of the mRNA would initiate at 3a, but a proportion would pass on to the next possible AUG, at the beginning of 3b, and those failing to initiate here would proceed to the start of 3c. Consistent with this possibility, the nucleotide sequence of 3a and 3b contains no AUG triplets, in any reading frame, other than those

linearized at different points. The plasmids were constructed as described under Methods and in Fig. 2. (b) Analysis of cell-free translation products of mRNA obtained by *in vitro* transcription from *Hin*dIII-digested pIBT4, pIBT5, pIBT6, pIBM6, *Pvu*II-digested pIBT4, and *Xba*I-digested pIBT4, using T7 RNA polymerase. RNA was added to the wheat germ cell-free system, as indicated above each lane, at approximately 100 μg/ml. Translation products were labeled with [^{35}S]methionine, separated on 22% SDS–polyacrylamide gel, and detected by fluorography. HMW and LMW, high- and low-molecular-weight markers, respectively. (c) Analysis of cell-free translation products of mRNA obtained by *in vitro* transcription from *Hin*dIII-digested pIBT1, pIBT2, and pIBT3, using T7 RNA polymerase.



Expected translation products			
pIBT12/HindIII	6.8K (3a)	7.4K (3b)	12.4K (3c)
pIBT13/HindIII	26.7K (3abc)	19.9K (3bc)	12.4K (3c)

FIG. 7. (a) Diagram of plasmids pIBT12 and pIBT13 showing the T7 promoter site, the 5′ UTR (constructed to correspond to the natural IBV mRNA 5′ leader sequence) in relation to the 3a, b, and c ORFs, and the restriction enzyme sites used to linearize the plasmids for *in vitro* transcription. Also shown are the sizes of the *in vitro* translation products expected from initiation at the beginning of each of the three ORFs on mRNA transcribed from plasmids linearized at different points. The plasmids were constructed as described under Methods. (b) Analysis of cell-free translation products of mRNA obtained by *in vitro* transcription from *Hin*dIII-digested pIBT4, pIBT12, pIBM6, and pIBT13, using T7 RNA polymerase. RNA was added to the wheat germ cell-free system, as indicated above each lane, at approximately 100 μg/ml. Translation products were labeled with [^{35}S]methionine, separated on a 22% SDS–polyacrylamide gel, and detected by fluorography. HMW, molecular weight markers.

used for initiation. However, in IBV-infected cells it appeared that the 3c polypeptide was more abundant than 3a or 3b, which would be surprising if 3c were only recognized by ribosomes which failed to initiate at 3a and 3b (particularly in view of our results indicating that the 3a initiation codon could function efficiently during *in vitro* translation). Direct comparison of the relative amounts of the 3a, 3b, and 3c polypeptides in infected cells, however, rests on the assumption that the antisera directed against 3a, 3b, and 3c recognize their targets with similar efficiencies; this appears to be the case using *in vitro*-synthesized proteins (data not shown), but may not be the case for the native proteins produced in infected cells. Nevertheless it remains quite possible that the 3c ORF may be translated from mRNA3 by some alternative to the leaky scanning mechanism. Synthetic tricistronic mRNAs derived from the plasmids pIBT3 and pIBT4 directed the synthesis of similar amounts of 3c *in vitro*, whereas production of 3a and 3b was very much lower from pIBT3 transcripts (Fig. 6), suggesting independent translational initiation of 3c. We are currently investigating this possibility further.

Other coronaviruses also produce mRNAs which appear to encode more than one protein. Sequence analysis has revealed that there are two overlapping ORFs (5a and 5b) with the potential to encode two polypeptides of 12.4kD (5a) and 10.2kD (5b) in the 5′ "unique" region of MHV mRNA5 (Skinner *et al.*, 1985), and the latter of these has been identified in MHV-infected cells (Liebowitz *et al.*, 1988). Comparison of the MHV mRNA5 with mRNA3 of IBV reveals several striking similarities. First, the 5′ "unique" region of each lies between the spike and the membrane genes. Second, in each case the downstream ORF located just before the start of the next subgenomic mRNA encodes a hydrophobic polypeptide with a stretch of nonpolar amino acids near its N-terminus (MHV 5b and IBV 3c). Third, the translation termination sites for these most distal ORFs appear variable among different strains of viruses. A deleted A residue at the end of the 5b ORF in the MHV strain A59 sequence should result in termination 5 amino acids earlier than in MHV-JHM. Likewise the 3c ORF of IBV strain 322/82 is truncated by 16 amino acids in comparison to the Beaudette strain, and the 68/84 and 183/66 strains also have different

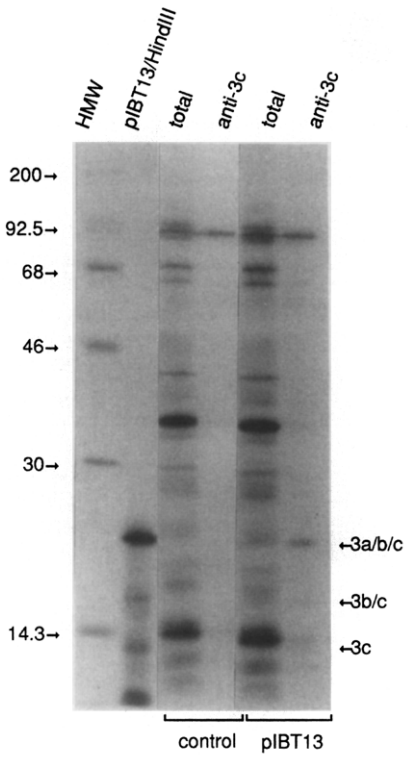


FIG. 8. Analysis of transiently expressed 3a, b, and c products from plasmid pIBT13 using the system described by Fuerst *et al.* (1986), in which CV1 cells were infected with a vaccinia recombinant virus expressing the T7 RNA polymerase and subsequently transfected with plasmid DNA. Cells were labeled with [^{35}S]methionine, lysates prepared, and polypeptides either analyzed directly or immunoprecipitated with the antisera indicated above each lane. HMW, molecular weight markers. The *in vitro* translation products obtained from RNA transcribed from pIBT13 (pIBT13/*Hin*dIII) were also included as a marker for the expected IBV specific products. Polypeptides were separated on a 17.5% SDS–polyacrylamide gel and detected by fluorography.

termination sites. Furthermore the ORFs immediately upstream of membrane protein-coding genes in other coronaviruses, such as transmissible gastroenteritis virus (Wesley *et al.*, 1989) and bovine coronavirus (Abraham *et al.*, 1990), have the potential to encode hydrophobic polypeptides with similarity to the 3c protein of IBV, suggesting that they could serve the same function in virus replication.

ACKNOWLEDGMENTS

This work was supported by a U.K. AFRC Link grant (LRG171) awarded to S.C.I., and by funds from the Ministry of Agriculture, Fishery and Food (D.C.). We thank Phil Davis and Rachel Purvis for excellent technical assistance. D. X. Liu is supported by a Cambridge Overseas Trust Studentship.

REFERENCES

ABRAHAM, S., KIENZLE, T. E., LAPPS, W. E., and BRIAN, D. A. (1990). Sequence and expression analysis of potential nonstructural proteins of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the spike and membrane protein genes of the bovine coronavirus. *Virology* **177,** 488–495.

BOURSNELL, M. E. G., BINNS, M., and BROWN, T. D. K. (1985). Sequencing of coronavirus IBV genomic RNA: Three open reading frames in the 5′ "unique" region of mRNA D. *J. Gen. Virol.* **66,** 2253–2258.

BOURSNELL, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. F., TOMLEY, F. M. and BINNS, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* **68,** 57–77.

BRIERLEY, I., BOURSNELL, M. E. G., BINNS, M. M., BILIMORIA, B., BLOK, V. C., BROWN, T. D. K., and INGLIS, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO. J.* **6,** 3779–3785.

BRIERLEY, I., DIGARD, P., and INGLIS, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57,** 537–547.

BROWN, T. D. K., BOURSNELL, M. E. G., and BINNS, M. M. (1984). A leader sequence is present on mRNA A of avian infectious bronchitis virus. *J. Gen. Virol.* **54,** 1437–1442.

BROWN, T. D. K., BOURSNELL, M. E. G., BINNS, M. M., and TOMLEY, F. M. (1986). Cloning and sequencing of 5′ terminal sequences from avian infectious bronchitis virus genomic RNA. *J. Gen. Virol.* **67,** 221–228.

CAVANAGH, D., and DAVIS, P. J. (1988). Evolution of avian coronavirus IBV: Sequence of the matrix glycoprotein gene and intergenic region of several serotypes. *J. Gen. Virol.* **69,** 621–629.

CAVANAGH, D., BRIAN, D. A., ENJUANES, L., HOLMES, K. V., LAI, M. M. C., LAUDE, H., SIDDELL, S. G., SPAAN, W., TAGUCHI, F., and TALBOT, P. J. (1990). Recommendations of the coronavirus study group for the nomenclature of the structural proteins, mRNAs, and genes of coronaviruses. *Virology* **176,** 306–307.

CHOMCZYNSKI, P., and SACCHI, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162,** 156–159.

CONTRERAS, R., CHEROUTRE, H., DEGRAVE, W., and FIERS, W. (1982). Simple efficient in vitro synthesis of capped RNA useful for direct expression of cloned DNA. *Nucleic Acids Res.* **10,** 6353–6362.

DAWSON, P. S., and GOUGH, R. E. (1971). Antigenic variation in strains of avian infectious bronchitis virus. *Archive fur die gesamte Virusforschung* **34,** 32–39.

DOTTO, G. P., ENEA, V., and ZINDER, N. D. (1981). Functional analysis of bacteriophage f1 intergenic region. *Virology* **114,** 463–473.

FUERST, T. R., NILES, E. G., STUDIER, F. W., and MOSS, B. (1986). Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **83,** 8122–8127.

INGLIS, S. C., CARROLL, A. R., LAMB, R. A., and MAHY, B. W. J. (1976). Polypeptides specified by the influenza virus genome. 1. Evidence for eight distinct gene products specified by fowl plague virus. *Virology* **74,** 489–503.

INGLIS, S. C., MCGEOCH, D. J., and MAHY, B. W. J. (1977). Polypeptides specified by the influenza virus genome. 2. Assignment of protein coding functions to individual genome segments by *in vitro* translation. *Virology* **78,** 522–536.

KOZAK, M. (1987). Analysis of 5′ noncoding sequences from 699 verbrate messenger RNAs. *Nucleic Acids Res.* **15,** 8125–8149.

KOZAK, M. (1989). The scanning model for translation: An update. *J. Mol. Biol.* **108,** 229–241.

KRIEG, P. A., and MELTON, D. A. (1984). Functional messenger RNAs are produced by SP6 in vitro transcription of cloned cDNA. *Nucleic Acids Res.* **12,** 7057–7071.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the bacteriophage T4. *Nature (London)* **227,** 680–685.

LAI, M. M. C. (1986). Coronavirus leader-RNA-primed transcription: An alternative mechanism to RNA splicing. *BioEssays* **5,** 257–260.

LEIBOWITZ, J. L., PERLMAN, S., WEINSTOCK, G., DEVRIES, J. R., BUDZILOWICZ, C., WEISSEMANN, J M., and WEISS, S. R. (1988). Detection of a murine coronavirus nonstructural protein encoded in a downstream open reading frame. *Virology* **164,** 156–164.

LOMNICZI, B. (1977). Biological properties of avian coronavirus RNA. *J. Gen. Virol.* **36,** 531–533.

RUSSEL, M., KIDD, S., and KELLEY, M. R. (1986). An improved filamentous helper phage for generating single-stranded plasmid DNA. *Gene* **45,** 333–338.

SCHOCHETMAN, G., STEVENS, R. H., and SIMPSON, R. W. (1977). Presence of infectious polyadenylated RNA in the coronavirus avian bronchitis virus. *Virology* **77,** 772–782.

SKINNER, M. A., EBNER, D., and SIDDELL, S. G. (1985). Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream, open reading frame. *J. Gen. Virol.* **66,** 581–592.

SMITH, A. R., BOURSNELL, M. E. G., BINNS, M. M., BROWN, T. D. K., and INGLIS, S. C. (1990). Identification of a new membrane-associated polypeptide specified by the coronavirus infectious bronchitis virus. *J. Gen. Virol.* **71,** 3–11.

STANLEY, K. K., and LUZIO, J. P. (1984). Construction of a new family of high efficiency bacterial expression vectors: Identification of cDNA clones coding for human liver proteins. *EMBO J.* **3,** 1429–1434.

STERN, D. F., and KENNEDY, S. I. T. (1980a). Coronavirus multiplication strategy. I. Identification and characterisation of virus specified RNA species to the genome. *J. Virol.* **34,** 665–674.

STERN, D. F., and KENNEDY, S. I. T. (1980b). Coronavirus multiplication strategy. II. Mapping the avian infectious bronchitis virus intracellular RNA species to the genome. *J. Virol.* **36,** 440–449.

STERN, D. F., and SEFTON, B. M. (1984). Coronavirus multiplication: Location of genes for virion proteins on the avian infectious bronchitis virus genome. *J. Virol.* **50,** 22–29.

WESLEY, R. D., CHEUNG, A. K., MICHAEL, D. D., and WOODS, R. D. (1989). Nucleotide sequence of coronavirus TGEV genomic RNA: Evidence for 3 mRNA species between the peplomer and matrix protein genes. *Virus Res.* **13,** 87–100.