Coronavirus IBV: partial amino terminal sequencing of spike polypeptide S2 identifies the sequence Arg-Arg-Phe-Arg-Arg at the cleavage site of the spike precursor propolypeptide of IBV strains Beaudette and M41

David Cavanagh¹, Philip J. Davis ¹, Darryl J.C. Pappin ², Matthew M. Binns ¹, Michael E.G. Boursnell ¹ and T. David K. Brown ¹

¹ Houghton Poultry Research Station, Houghton, Huntingdon, Cambs. PE17 2DA, and ² Department of Biochemistry, University of Leeds, Leeds LS2 9JT, U.K.

(Accepted for publication 4 October 1985)

Summary

The spike protein of avian infectious bronchitis coronavirus comprises two glycopolypeptides S1 and S2 derived by cleavage of a proglycopolypeptide So, the nucleotide sequence of which has recently been determined for the Beaudette strain (Binns, M.M. et al., 1985, J. Gen. Virol. 66, 719–726). The order of the two glycopolypeptides within So is aminoterminus(N)-S1-S2-carboxyterminus(C). To locate the N-terminus of S2 we have performed partial amino acid sequencing on S2 from IBV-Beaudette labelled with [¹³H]serine and from the related strain IBV-M41 labelled with [³H]valine, leucine and isoleucine. The residues identified and their positions relative to the N-terminus of S2 were: serine, 13; valine, 6, 12; leucine, none in the first 20 residues; isoleucine, 2, 19. These results identified the N-terminus of S2 of IBV-Beaudette as serine, 520 residues from the N-terminus of S1, excluding the signal sequence. Immediately to the N-terminal side of residue 520 So has the sequence Arg-Arg-Phe-Arg-Arg; similar basic connecting peptides are a feature of several other virus spike glycoproteins. It was deduced that for IBV-Beaudette S1 comprises 519 residues (M, 57.0K) or 514 residues (56.2K) if the connecting peptide was to be removed by carboxypeptidase-like activity in vivo while S2 has 625 residues (69.2K). Nucleotide sequencing of the cleavage region of the So gene of IBV-M41 revealed the same connecting peptide as IBV-Beaudette and that the first 20 N-terminal residues of S2 of IBV-M41 were identical to those of the Beaudette strain. IBV-Beaudette grown in Vero cells had some uncleaved So; this was cleavable by 10 μg/ml of trypsin and of chymotrypsin.
Partial N-terminal analysis of S1 from IBV-M41 identified leucine and valine residues at positions 2 and 9 respectively from the N-terminus. This confirms the identification, made by Binns et al. (1985), of the N-terminus of S1 and the end of the signal sequence of the IBV-Beaudette spike propolypeptide.

N-terminal sequencing of [³H]leucine-labelled IBV-Beaudette membrane (M) polypeptide showed leucine residues at positions 8, 16 and 22 from the N-terminus; these results confirm the open reading frame identified by M.E.G. Boursnell et al. (1984, Virus Res. 1, 303–313) in the nucleotide sequence of M. The N-terminus of the nucleocapsid (N) polypeptide appeared to be blocked.

coronavirus IBV, protein sequencing, spike, propolypeptide, cleavage

Introduction

Avian infectious bronchitis virus (IBV), a pathogen of chickens, belongs to the genus coronavirus. These enveloped viruses have a nucleocapsid protein (N; \( M_r \approx 50K \)) associated with the single-stranded plus-sense RNA genome and two glycoproteins embedded in the membrane (for review see Siddell et al., 1983). The larger of the two glycoproteins is the spike (S) or peplomer protein which in IBV has two or three copies of each of glycopolypeptides S1 (\( M_r \approx 90K \)) and S2 (\( \approx 84K \)) (Cavanagh, 1983a, b). One function of S2 is to anchor the spike in the virus membrane (Cavanagh, 1983b). S1 and S2 are derived by cleavage of a precursor glycopolypeptide So (Stern and Sefton, 1982). The amino acid sequence of So of the Beaudette strain of IBV has been deduced by nucleotide sequencing (Binns et al., 1985). Amino acid sequencing of the amino- (N-) terminus of [³H]serine labelled S1 showed that S1 started immediately after the putative signal peptide at the N-terminus of So. This indicated that S2 was contained within the C-terminal half of So and that the N-terminus of S2 was generated when So was cleaved near the middle of the molecule. This paper described partial amino acid sequencing of S2 from two serologically related strains of IBV and correlation with the nucleotide sequence of So in order to identify the propolypeptide cleavage site.

Methods

Virus strains

The Beaudette strain of IBV that was used for N-terminal sequencing was obtained from Dr. Bela Lomniczi and was subsequently passaged six times in primary chick kidney (CK) cells prior to plaque purification three times in CK cells. Vero cell-adapted IBV-Beaudette (Alonso-Caplen et al., 1984) was given by Dr. B. Mahy who obtained it from Dr. R. Compans. After four additional passages in Vero cells (ATCC number CCL 81; Flow Laboratories, Scotland, U.K.) the virus was passaged once in 11-day-old embryonated chicken eggs. The allantoic fluid was
harvested when some of the embryos had died; the virus titre was $7.3 \log_{10} 50\%$ egg lethal doses. IBV-M41 was grown in embryonated eggs.

**Preparation of radiolabelled virus**

For N-terminal sequencing IBV-Beaudette was radiolabelled in CK cells (Stern and Sefton, 1982) and IBV-M41 in de-embryonated eggs (Cavanagh, 1981). For labelling in Vero cells the cells were inoculated with undiluted allantoic fluid containing $7.3 \log_{10} 50\%$ egg lethal doses. After 2 h in air at $37^\circ C$ the procedure used was essentially that described by Stern and Sefton (1982) for CK cells except that virus was harvested at 28–36 h after infection. Virus was labelled with $1 \mu$Ci of $[3,3^\text{H}]$serine, $[3,4(\eta)-3^\text{H}]$valine, $[4,5-3^\text{H}]$leucine or $[4,5-3^\text{H}]$isoleucine plus $0.1 \mu$Ci of $[^{35}\text{S}]$methionine, all from Amersham International. Virus was purified by sucrose gradient centrifugation (Cavanagh, 1981).

**SDS–polyacrylamide gel electrophoresis (SDS–PAGE), electroelution and N-terminal sequencing**

SDS–PAGE and electroelution have been described (Cavanagh, 1983b; Binns et al., 1985) except that the polypeptides containing $^3\text{H}$-labelled amino acids plus $[^{35}\text{S}]$methionine were located after SDS–PAGE by autoradiography and not fluorography. Electroeluted, lyophilised polypeptides were dissolved in $0.07 \text{ ml}$ of $0.2 \text{ M}$ sodium bicarbonate containing $0.25\%$ SDS and added to $20 \text{ mg}$ dry $p$-phenylendioisothiocyanate (DITC) glass ($170 \text{ Å}$ pore-size, 200–400 mesh) prepared according to the method of Wachter et al. (1973). The glass/polypeptide suspension was incubated at $56^\circ C$ for 90 min under nitrogen and the glass then washed extensively with water and methanol to remove noncovalently bound material. Attachment yields were monitored by radioactivity. The glass-coupled polypeptide was then sequenced by automated solid-phase Edman degradation (Laursen, 1971) using the microsequence solid-phase facility built by the Agricultural and Food Research Council Sequencing Unit, Department of Biochemistry, University of Leeds. The normal 30 min sequence cycle (at $56^\circ C$) comprised 18 min coupling with $5\%$ (v/v) phenyl isothiocyanate (PITC) in sequence buffer (Pappin and Findlay, 1984), 3 min wash with methanol (2 ml/min), 3 min wash with benzene (2 ml/min), followed by 5 min cleavage with anhydrous trifluoroacetic acid (TFA) (0.2 ml/min). TFA fractions containing the cleaved anilinothiazolinone amino acids from each sequence cycle were collected, the acid removed by evaporation in vacuo over NaOH flake, and the whole sample counted in 4.5 ml ‘Liquiscint’ scintillation cocktail.

**Protease treatment of virus**

Virus from sucrose gradients was incubated at $37^\circ C$ with either TPCK-treated trypsin (Sigma type XIII, from bovine pancreas) or TLCK-treated chymotrypsin (Sigma type VII, from bovine pancreas). After 30 min the trypsin inhibitor TLCK (Sigma) was added (500 $\mu$g/ml) to trypsin-containing samples, followed by addition to all samples of SDS (2%) and mercaptoethanol (2%) and heating at $100^\circ C$ for 2 min.
Cloning and sequencing of the So gene of IBV-M41

The cDNA cloning of sequences encoding the M41 spike protein was as described previously (Binns et al., 1985), except for the use of a 15-base oligonucleotide primer complementary to a sequence in the genomic RNA present towards the putative 5' end of the body of mRNA D. cDNA clones containing viral inserts were identified by colony hybridisation using 32P-end-labelled fragmented genomic virus RNA as a probe. Clone pMB233 was characterised further and found to extend 5'-wards from the primer for approximately 2200 base pairs. M13/dideoxynucleotide sequencing (Sanger et al., 1977; Biggin et al., 1983) of the S1/S2 junction was carried out on PstI fragments of pMB233 which had been subcloned into PstI-digested M13mp10.

Results

N-terminal sequence analysis of the M and N polypeptides

Analysis of [3H]leucine- and [3H]serine-labelled M from IBV-Beaudette identified leucine residues at positions 8, 16 and 22 and a serine residue at position 13 from the N-terminus (Fig. 1, A, B). The relatively high amount of radiolabel after cycles one and two (Fig. 1, B) is an artefact. This arose because the time allowed for acid cleavage of the first residue, proline, was doubled to increase the sequencing efficacy through this residue. This results in a greater than normal leaching of the polypeptide from the glass support. To avoid this in subsequent analyses the cleavage times at proline residues was not increased.

When [3H]leucine- and [3H]serine-labelled N polypeptides from IBV-Beaudette were examined no labelled residues were detected even though more than 500000 dpm were covalently attached to the support.

N-terminal sequence analysis of the S1 and S2 polypeptides

Analysis of [3H]valine- and [3H]leucine-labelled S1 from IBV-M41 showed valine and leucine residues at positions 9 and 2 respectively (Fig. 1, C, D).

Sequencing of S2 from IBV-Beaudette labelled with [3H]serine identified one serine residue in the first 20 residues, at position 13 (Fig. 1, E). S2 from IBV-M41 was labelled with [3H]valine, [3H]leucine and [3H]isoleucine for further analyses. Valine residues were detected at positions 6 and 12 (Fig. 1, F), isoleucine residues at positions 2 and 19 (Fig. 1, G) and no leucine residues in the first 20 residues analysed (Fig. 1H).

Amino acid sequence of the cleavage region of So of IBV-M41

A fragment of clone pMB233 was used for sequencing the cleavage region of the S gene of IBV-M41. Fig. 2B shows the amino acid sequence deduced from the nucleotide sequence in this region, which was identical to the equivalent sequence in the S1/S2 cleavage region of the IBV-Beaudette S gene.

Cleavage of So by trypsin and chymotrypsin

IBV-Beaudette which had been adapted to growth in Vero cells (Alonso-Caplen et
Fig. 1. Partial N-terminal amino acid sequencing of IBV polypeptides. Radiolabelled polypeptides were covalently attached to 170 Å DITC-glass and subjected to solid-phase Edman degradation. Attachment yields to the glass support were better than 90% in all cases: (A, B) M from IBV Beaudette labelled with (A) [3H]leucine and (B) [3H]serine; (C, D) S1 from IBV-M41 labelled with (C) [3H]valine and (D) [3H]leucine; (E) S2 from IBV-Beaudette labelled with [3H]serine; (F–H) S2 from IBV-M41 labelled with (F) [3H]valine, (G) [3H]isoleucine and (H) [3H]leucine. Owing to the nature of the DITC-attachment procedure, residue one cannot be identified as it remains covalently attached to the glass support following TFA cleavage. Repetitive sequence yields estimated for the S2 polypeptide were as follows (measured between counts recovered at the relevant cycles minus the mean background): Ile 2-Ile 19, 94.9%; Val 6-Val 12, 93.9%.
Fig. 2. (A) Part of the amino acid sequence of the spike precursor polypeptide of IBV-Beaudette, deduced from nucleotide sequence data (Binns et al., 1985). The sequence shown starts at the beginning of the signal sequence (underlined). The numbers commence at the N-terminal residue (valine) of the mature S1 polypeptide, i.e. after removal of the signal sequence. Asterisks (*) mark the leucine (L) and valine (V) residues identified by partial amino-terminal sequencing. The valine residue at position one could not be detected by the technique used (see legend to Fig. 1). (B) Part of the deduced amino acid sequence of the spike precursor polypeptide of IBV-M41; IBV-Beaudette has an identical sequence. The numbers below the sequence correspond to the residue positions from the N-terminus of the spike precursor, excluding the signal sequence, of IBV Beaudette. The numbers above the sequence start at the N terminal residue (serine) of S2. Asterisks (*) mark the serine (S), isoleucine (I) and valine (V) residues identified by partial amino acid sequencing. The proposed connecting peptide is underlined. A potential glycosylation site for an N-linked glycan is indicated by \(\oplus\) over the asparagine residue.

al., 1984) was radiolabelled with \(^{35}\)S]methionine. SDS–PAGE revealed five virus-specific polypeptides (Fig. 3, track a). Minor polypeptides detected in some gels were considered to be host components since SDS–PAGE analysis of every fraction of the preparative gradient showed that the peaks of these minor bands was not coincident with that of the known viral polypeptides. When IBV-Beaudette is grown in chick kidney cells the resultant virus does not contain So since all is cleaved to S1 and S2 (Stern and Sefton, 1982; Cavanagh and Davis, unpublished observation). However, virus grown in Vero cells had, in addition to S1 (\(M_r\) 91K) and S2 (86K), some So of \(M_r\) 159K; this is similar to the \(M_r\) estimate of 155K for cell-associated So (Stern and Sefton, 1982). The N and M polypeptides were estimated to have \(M_r\) s of 55K and 31K. Trypsin (10 \(\mu\)g/ml, 30 min, 37°C) cleaved So (Fig. 3, track b). No other
polypeptide was hydrolysed, even with enzyme at 1 mg/ml. Chymotrypsin (10 
μg/ml, 30 min, 37°C) also hydrolysed So (Fig. 3, track f).

Discussion

Nucleotide sequence analysis of the M gene of IBV-Beaudette identified a long open reading frame predicting a polypeptide of 225 amino acids (Boursnell et al., 1984). Partial N-terminal analysis of M has now confirmed the choice of open reading frame. Moreover, it shows that the N-terminal methionyl residue specified by the first codon of the gene is not present in virion-associated M.
N-terminal sequencing of N labelled with \[^{3}H\]\text{leucine} and \[^{3}H\]\text{serine} failed to reveal either residue in the first 20 amino acids, although nucleotide sequence data had indicated that both of these do occur within the first 20 residues (Boursnell et al., 1985). This indicates either that the N-terminus of N is blocked or that the open reading frame identified by nucleotide sequencing is incorrect. The latter is unlikely, however, since a 1227-base open reading frame was identified within the N gene of both IBV-M41 and IBV-Beaudette, with only a 5.6% difference in amino acids between them and with considerable homology with the sequence of the N poly-peptide of murine hepatitis coronavirus. Thus, the N-terminus of N is probably blocked.

N-terminal sequence analysis of \[^{3}H\]\text{serine}-labelled S1 from IBV-Beaudette showed that serine residues were present at positions 5, 6, 7, 14 and 20 from the N-terminus and that the N-terminal residue of mature S1 was at position 19 from the N-terminus of So i.e. the signal peptide was 18 residues long (Binns et al., 1985; Fig. 2A). Analysis of the first 20 residues of \[^{3}H\]\text{valine}-labelled S1 from IBV-M41 showed that a valine residue was at position 9. Inspection of the deduced sequence of IBV-Beaudette S1 indicates that there are two valine residues (Fig. 2A). One of these is residue 9 while the other is residue 1. The latter residue would not have been detected by amino acid sequencing since it would have been covalently bound to the glass support and hence not released during Edman degradation. Amino acid analysis of \[^{3}H\]\text{leucine}-labelled S1 from IBV-M41 showed a leucine residue at position 2 and no others in the first 20 residues. This agrees with the deduced IBV-Beaudette sequence (Fig. 2A) and confirms the choice of open reading frame for S1 of IBV-Beaudette and indicates that there is some similarity between S1 of IBV strains M41 and Beaudette within the first 20 N-terminal residues.

The amino acid sequence of residues 514–539 of So of IBV-Beaudette (Binns et al., 1985) is shown in Fig. 2B. This sequence was also found to be present within So of IBV-M41, as determined by nucleotide sequencing of a cDNA clone of IBV-M41. The positions, with respect to the N-terminus, of the residues identified by partial N-terminal sequencing of S2 from IBV-Beaudette and M41 were: serine (S), 13; valine (V), 6 and 12; isoleucine (I), 2 and 19; and no leucine (L) residues in the first 20 residues. This data unequivocally identifies the N-terminal residue of S2 of IBV-Beaudette as serine residue 520 from the N-terminal of So (excluding the signal peptide) (Fig. 2B).

Cleavage of the IBV-Beaudette So between residues 519 and 520 would generate an S1 of 519 residues with an $M_r$ of 57 000. However, if the basic residues of the cleavage site are removed by carboxypeptidase B type activity, as occurs with influenza virus haemagglutinin (HA) (Garten et al., 1981), then S1 would have 514 residues of $M_r$ 56 228. In either event, S2 is predicted to have 625 residues of $M_r$ 69 208. Thus the polypeptide moiety of S2 has an $M_r$ greater than S1, whereas the reverse is the case for the glycosylated molecules. One factor that may contribute to this difference is that S1 has four more potential glycosylation sites than S2 (Binns et al., 1985).

The sequence Arg-Arg-Phe-Arg-Arg (RRFRR in single letter code) which precedes the N-terminal serine residue of S2 is very interesting (Fig. 2B). It resembles
the basic sequences which form the cleavage sites of glycoproteins of several other enveloped viruses. Thus two alphaviruses have Arg-His-Arg-Arg and another has Arg-Ser-Lys-Arg preceding the N-terminal serine residue of the E2 glycopolypeptide (Garoff et al., 1980; Rice and Strauss, 1981; Dalgarno et al., 1983). Two retroviruses have Arg-Arg-Lys-Arg or Arg-His-Lys-Arg at the env propolypeptide (precursor polypeptide) cleavage site (Shinnick et al., 1981; Schwartz et al., 1983). Paramyxovirus SV5 (Paterson et al., 1984) and orthomyxovirus fowl plague virus (Porter et al., 1979) have cleavage sequences Arg-Arg-Arg-Arg-Arg and Lys-Lys-Arg-Glu-Lys-Arg in the fusion propolypeptide and haemagglutinin propolypeptide, respectively. Many propolypeptides of polypeptide hormones have two basic residues at the post-translational cleavage site (Docherty and Steiner, 1982). One type of enzyme involved in the cleavage of propolypeptides has trypsin-like activity (Docherty and Steiner, 1982). We have shown that S0 of IBV can be cleaved by trypsin, as can the spike precursor of MHV (Sturman and Holmes, 1977). Thus, the process by which S0 is cleaved to yield S1 and S2 probably resembles that of many other propolypeptides. It would be expected that a sequence as basic as the Arg-Arg-Phe-Arg-Arg cleavage site of IBV-Beaudette would be at the surface of the spike protein. The phenylalanine residue in the middle of the cleavage peptide is a potential target residue for chymotrypsin. In view of the likely exposed position of the cleavage peptide it would be expected that S0 would be cleaved by chymotrypsin; this was the case. Whether any of the basic residues are removed by carboxypeptidase-like activity in vivo (Docherty and Steiner, 1982) remains to be determined.

The amino terminus of HA2 of influenza virus which is generated by cleavage of HAo is hydrophobic and its sequence is conserved among type A and B influenza viruses (Skehel and Waterfield, 1975; Gething et al., 1980) and has some homology with the hydrophobic N-terminus of F1 derived by cleavage of the fusion propolypeptide of orthomyxoviruses (Gething et al., 1978). The N-terminus of S2 is hydrophobic although not more so than several other regions of S0. Whether the N-terminus of S2 has an important role in membrane fusion, as is the case with the N-terminus of HA1 and F1 (for a review, see White et al., 1983) is not known although cleavage of the MHV spike precursor is necessary for fusion-from-without (Sturman and Holmes, 1984). The cleavage of the orthomyxovirus haemagglutinin propolypeptide (HAo) and the paramyxovirus fusion propolypeptide is essential for full biological activity (Klenk and Rott, 1980). Cleavage enables efficient fusion of the virus membrane with the plasma or endosome membrane to occur; the RNA genome then enters the cytoplasm.

Whether the host-cell range of IBV strains correlates in part with the number of basic residues at the S0 cleavage site, as is the case for orthomyxoviruses (Bosch et al., 1979, 1981), is under investigation.

Acknowledgements

We thank Bridgette Britton and Judy Thompson for excellent technical assistance. This work was supported in part by Research Contract No. GB1-2-011-UK of
the Biomolecular Engineering Programme of the Commission of the European Communities.

References


(Manuscript received 19 July 1985)