Volume 14 Number 12 1986 Nucleic Acids Research

Genomic structure of the large RNA segment of infectious bursal disease virus

Peter J.Hudson, Neil M.McKern, Barbara E.Power and Ahmed A.Azad

CSIRO, Division of Protein Chemistry, Parkville 3052, Australia

Received 8 April 1986; Accepted 13 May 1986

ABSTRACT

The larger RNA segment of infectious bursal disease virus (IBDV:Australian strain 002-73) has been characterized by cDNA cloning and nucleotide sequence analysis. We believe IBDV is the first birnavirus to be sequenced and so have confirmed the coding region by N-terminal amino acid sequence analysis of intact viral proteins and several tryptic peptide fragments. The large RNA segment encodes in order the 37-kDa, 28-kDa and 32-kDa proteins within a continuous open reading frame and the primary translation product appears to be subsequently processed into the mature viral proteins. The large protein precursor is still processed into the 32-kDa host protective immunogen when expressed as a fusion protein in E. coli. These results are in marked contrast to the predictions from in vitro translation data that birnavirus genomes are expressed as polycistronic templates. We can now propose that birnaviruses, in particular IBDV, possess monocistronic segments and that the precursor is proteolytically processed in vivo. The sequence data presented for the 32-kDa host protective immunogen may provide the basic information needed for the production of an effective subunit vaccine against this commercially important virus.

INTRODUCTION

Infectious bursal disease virus (IBDV) can establish a highly contagious disease in young chickens that causes severe immunodepression due to destruction of developing β-lymphocytes in the bursa of Fabricus (1). Australian strain (002-73) as well as overseas strains of the virus, are capable of causing significant commercial losses to the poultry industry (2,3). The current methods for producing an inactivated vaccine are expensive, labour intensive and subject to considerable batch variation (3,4). We have characterized the IBDV genome as the first step towards the production of and effective subunit vaccine for the control of this commercially important disease.

The genome of the Australian IBDV strain (002-73) consists of two segments of double-stranded RNA (5) typical of birnaviruses (6).

The smaller RNA segment in IBDV (about 2900 bp) encodes the 90-kDa viral RNA polymerase and is not discussed further in this report. The larger RNA segment comprises about 3400 bp and has been shown to encode three viral proteins; the 41-kDa precursor to the 37-kDa structural protein, a 28-kDa protein and the 32-kDa structural protein (5). Of these, the 32-kDa antigen is the most likely target for the production of a subunit vaccine since the purified protein when injected into chickens produces antibodies capable of both neutralizing the virus in vitro and affording passive protection (7). A cDNA library has been prepared which contains fragments spanning the majority of the larger IBDV RNA segment (5). These cDNA clones have been used as the starting material for our analysis of the IBDV genome.

METHODS

Molecular Cloning of the IBDV Genome

Techniques for colony hybridization, isolation of plasmid DNA, production of hybridization probes, and autoradiography were performed as described earlier (5). A recombinant cDNA library has been prepared from IBDV genomic RNA into the plasmid vector pBR322 (5). A mixed population of cDNA fragments (350-2000 bp) spanning the large RNA segment of the IBDV genome were then subcloned, still containing homopolymeric dG/dC tails, into pUR expression vectors (8) and screened with antibodies specific for the 32-kDa protein as described (9).

Nucleotide Sequence Analysis

Modifications were made to both the chemical degradation technique (10) and the dideoxy chain termination method (11). Internal cDNA sequences in either pBR322 or pUR vectors were obtained by a cleavage and end-labelling strategy at a number of suitable restriction sites chosen for their ability to incorporate α-^{32}P-nucleotides (12). In the pUR expression plasmids sequencing by this method from the EcoRI site enabled the nucleotide sequence to be obtained directly over the dG/dC tails into the cDNA structure and allowed the reading phase to be determined in either pUR 290, 291 or 292. As an alternative method the dG/dC homopolymeric tails were removed from cDNA inserts using Bal 31 exonuclease in a controlled reaction (2 units, 20°C, 10 mins) designed to digest no more than 50 nucleotides from either end. Fragments were then blunt-ended with DNA polymerase I (Klenow fragment) and ligated into a Sma I restricted M13mp10 vector followed by transformation of

E. coli JM101 cells (11). Single-stranded templates were sequenced by the primed synthesis method using an M13-specific primer (11) but with modifications that improved transcription fidelity over regions of secondary structure in the template. These included removal of NaCl from the buffer, using reverse transcriptase and optimized ratios of dideoxy:deoxy-nucleotides (1:30A; 2:15C; 1:15G; 2:3T) and performing the reaction at 30°C or greater. When necessary sequencing samples from any of the above methods were loaded on denaturing polyacrylamide gels (11) which contained 90% formamide. Under these conditions, when 20 cm x 40 cm gels were run at 25W on an apparatus that maintained the temperature above 50°C, the secondary structure was completely disrupted. Nucleotide sequences were compiled using a VAX/VMS computer system using the programmes described previously (13) with modifications by Dr. T. Kyne (unpublished).

Amino Acid Sequence Analysis

The IBDV proteins were excised and electroeluted from SDS-polyacrylamide gels by the method described in (14). Purified protein was then digested with 2% (w/w) TPCK trypsin for 4 h at 37°C in 0.05M NH_4HCO_3, the solution dried and the residue redissolved in 0.1% trifluoroacetic acid. Soluble material was injected onto a Vydac 218TP 5μ reverse phase column and digest products were eluted as described elsewhere (15) by high pressure liquid chromatography. Peptides were manually sequenced (16) using previously described modifications to the Edman degradation technique (15).

Analysis of Expressed Proteins

IBDV cDNA fragments were subcloned into pEX vectors (17) for expression in E. coli cells. Recombinant clones were grown in liquid media containing 50μg/ml ampicillin at 30°C to an O.D. 650 nm of 0.1 and induced at 42°C for four hours (9). Cells were disrupted with 2% SDS and subjected to polyacrylamide gel electrophoresis before Western blotting onto nitrocellulose and probing with monoclonal antibodies as described (9).

RESULTS AND DISCUSSION

Characterization of the Large RNA Segment of IBDV

An IBDV cDNA library had been constructed in pBR322 (5) and shown to contain over one hundred unique recombinants with overlapping cDNA inserts that spanned the large RNA segment. The cDNA map of the large

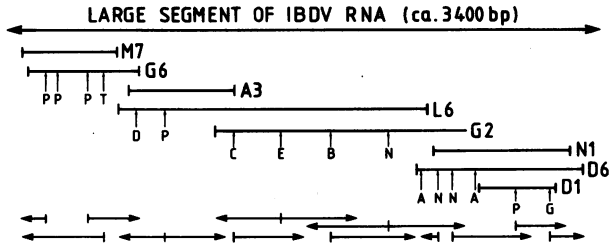

Figure 1

Nucleotide sequencing strategy and map of the large RNA segment of IBDV. Overlapping cDNA clones designated M7 to D6 have been previously mapped by cross-hybridization (5). Chemical sequencing (10) was performed on cDNA fragments end-labelled at suitable restriction sites and the sequence runs are shown as arrowed lines. The restriction sites used for sequence analysis are shown as Apa I (A), Bam HI (B), Sca I (C), Nde I (D), Eco RI (E), Bgl II (G), Hind III (H), Nco I (N), Hpa II (P) and Taq I (T). Over 60 unique dideoxy sequencing (11) runs were performed on Bal 31 exonuclease-treated cDNA fragments subcloned into M13 in order to confirm overlaps and complete the nucleotide sequence (data not shown).

RNA segment (5) together with a summarized sequencing strategy is presented in Fig. 1. The nucleotide sequence was compiled by dideoxy sequencing on at least 60 independent cDNA clones as described in Methods. Regions of ambiguity were resolved by an end-labelling and chemical degradation strategy from suitable restriction sites. In the final version of the nucleotide sequence presented in Fig. 2 each residue had been confirmed from at least two overlapping sequencing runs. The strategy of sequencing independent cDNA clones provided a check on the fidelity of cDNA synthesis during the construction and amplification of the cDNA libraries. There were no point mutations, deletions or rearrangements found in the cDNA inserts thus confirming that the method of random-primed cDNA synthesis (5) was remarkably error-free.

The only serious problem encountered was that of poor quality nucleotide sequence obtained by the chain termination method (11) due to multiple regions of secondary structure that caused premature terminations. These regions were resolved by the use of reverse transcriptase under optimized conditions rather than the standard reaction with DNA polymerase (Klenow fragment). Occasionally it was necessary to use the chemical degradation technique (10) which is less affected by secondary structure and the use of denaturing formamide gels to resolve compressions on the sequencing ladder. It is possible that

the IBDV genome, being double stranded RNA, can tolerate a high degree of secondary structure without affecting transcription or replication. However, the difficulties reported during the construction of the original cDNA library (5) may have been due to this secondary structure. Furthermore, we were unable to extend the sequence analysis to the ends of the gene segment by the use of synthetic oligonucleotide primers probably due to the same secondary structure problems.

Prediction of the Precursor Structure

A single continuous open reading frame could be translated from the nucleotide sequence obtained for the large RNA segment (Fig. 2). This result was rather surprising given previous interpretations of *in vitro* translation data (5,18) which had suggested that the birnavirus large RNA segment is a polycistronic template. We found only a single termination codon which presumably defines the carboxyl terminus of the precursor at residue 1011 (Fig. 2). To confirm the predicted monocistronic reading frame we have analysed both the amino acid sequence of native viral proteins and the nucleotide sequence of *E. coli* expression clones.

The 37-kDa, 28-kDa and 32-kDa proteins were isolated by SDS-polyacrylamide gel electrophoresis and subjected to Edman degradation as described in Methods. The N-terminus of the 37-kDa was identified as threonine (residue 1 in Fig. 2). However, the N-termini of the 28-kDa and 32-kDa proteins were blocked and it was therefore necessary to isolate and sequence several tryptic peptides. The results outlined in Figure 2 confirm both the coding region assignment and the predicted reading phase. The IBDV large RNA segment encodes in order the 37-kDa, 28-kDa and 32-kDa viral proteins within a continuous open reading frame.

As further confirmation of the reading frame we analysed two clones (D6 and D1) which were described previously as expressing regions of the 32-kDa antigen fused to β-galactosidase (9). From the nucleotide sequence the fusion proteins were shown to consist of an IBDV-encoded peptide (Fig. 2) fused to the C-terminus of β-galactosidase via a polyglycine segment derived from the homopolymeric dG/dC tails. The nucleotide sequence also established that both the D6 and D1 clones expressed fusion proteins in the same translation reading phase as the open reading frame in Figure 2. The 310 residue peptide encoded by the cDNA insert in D6 (residue 702-1011) had a predicted molecular weight of

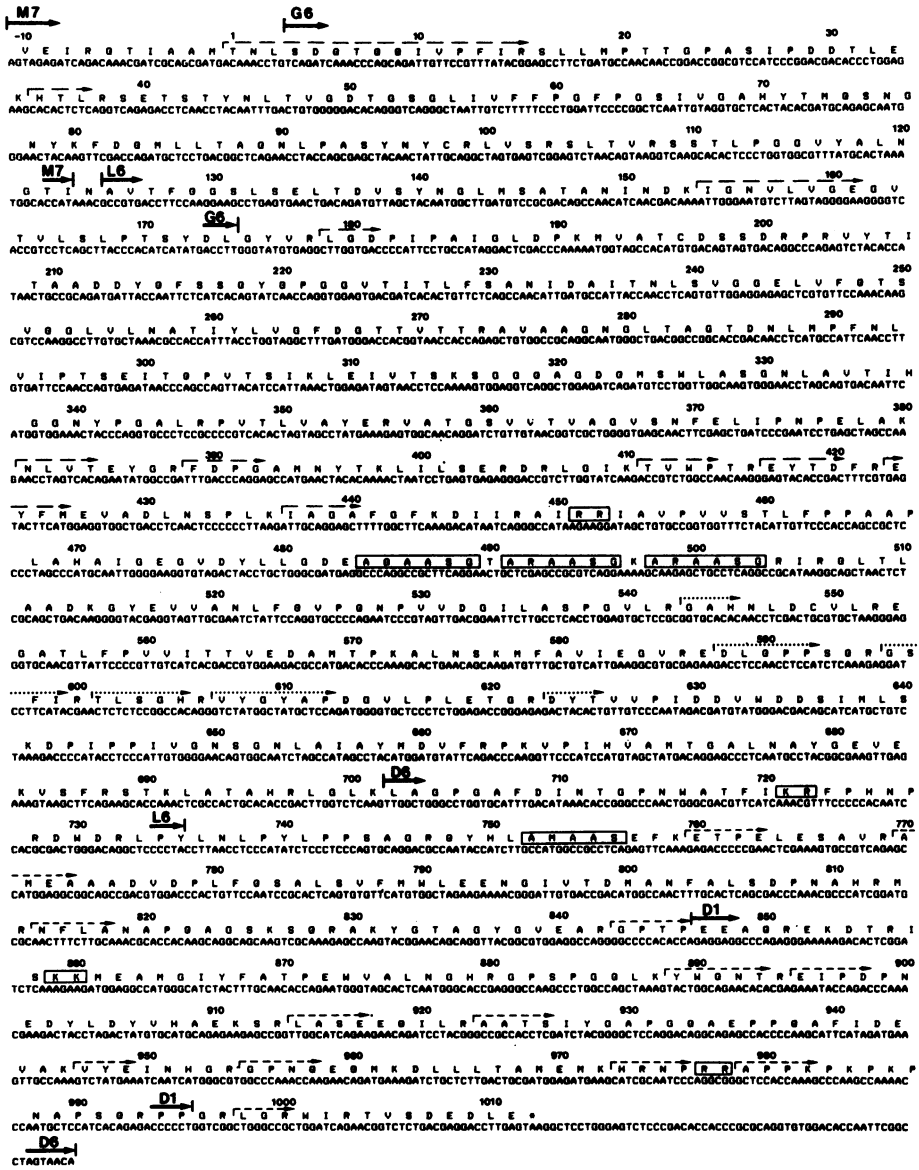

Figure 2

Sequence analysis of the large RNA segment of IBDV. The predicted amino acid sequence is presented in single letter code above the nucleotide sequence derived from cDNA clones. There are no other extensive open reading frames. The amino acid sequences are numbered sequentially from the N-terminus of the 37-kDa protein as position 1. The region encompassed by cDNA clones M7, G6, L6, D6 and D1 are indicated. Dibasic residues are boxed and the repeat

unit A-X-A-A-S is similarly highlighted. N-terminal sequences derived from tryptic peptides are shown overlined as (—— ——→) for the 37-kDa, (••••>) for 28-kDa and (----->) for the 32-kDa protein. Only the N-terminus of the 37-kDa protein could be obtained by direct sequencing on intact proteins and this is shown from residue 1.

34.7-kDa and ended with the IBDV encoded termination codon. Since the tryptic peptide sequence data confirmed the IBDV 32-kDa antigen spanned at least residues 760-1000 it is likely that the D6 clone expresses most, if not all of the native 32-kDa antigen.

Processing of the Precursor in vivo

To account for the incorrect predictions of a polycistronic template over the large RNA segment by in vitro translation studies (5,18) it was likely that the precursor was processed either by viral encoded or cellular proteases. We therefore constructed a vector (designated PO) capable of expressing most of the large RNA segment (residues 5-1011) in E. coli. The IBDV genome was reconstructed by the ligation of a Sau 3a-Nde I fragment of the cDNA clone G6 (encoding residues 5-171) with an Nde I-Apa I fragment of L6 (residues 172-713) and an Apa I-Pst I fragment of D6 (residues 714-1011) into a Bam HI-Pst I digested pEX3 vector. The cDNA reconstruction of the large RNA segment was checked by restriction mapping analysis and the reading phase over the fusion region was confirmed by nucleotide sequencing.

The pUR and pEX expression vectors have been designed for ligation of cDNA fragments into the 3'end of the β-galactosidase gene in order to produce stable fusion proteins (8,17). We have previously described the expression of cDNA clones D6 and D1 (Fig. 2) in such vectors and have shown that the fusion protein appears as a single protein band of approximately 140 and 120-kDa respectively by western gel analysis (9). Clone PO, in addition to producing the expected fusion protein, also expressed a 32-kDa protein as the major product recognised by monoclonal antibodies specific for the 32-kDa antigen (Figure 3). The expressed 32-kDa protein is evidently a stable product of the processed precursor in E. coli and appears to be similar to the native viral 32-kDa protein since they co-migrate on SDS-polyacrylamide gels (data not shown). Unfortunately we do not yet have available monoclonal antibodies which recognise the 37-kDa or 28-kDa proteins and so cannot determine the fate of these proteolytic fragments.

The processing of in vitro translation products has recently been

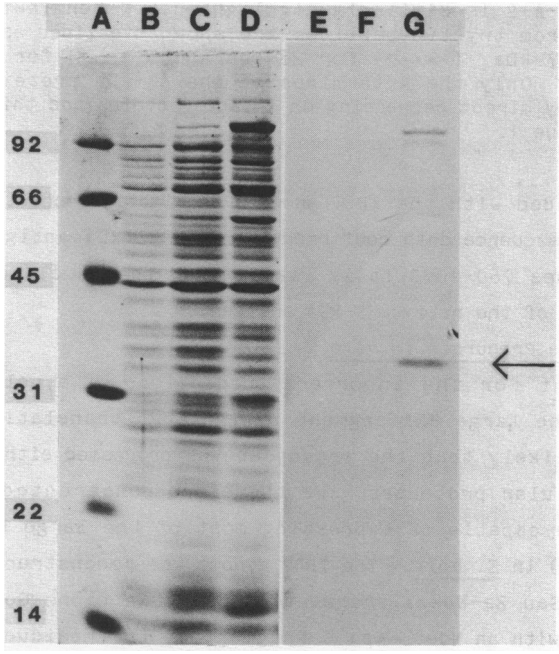

Figure 3
Analysis on SDS-polyacrylamide gels of the expressed IBDV precursor (residues 5-1011) fused to β-galactosidase in the pEX2 vector (17). Lanes a-d are Coomassie stained tracks of total E. coli cell extracts and lanes e-g are the corresponding Western transfers of a similar SDS-gel visualized with monoclonal antibodies specific for the 32-kDa antigen as described (9). Lane a, standard molecular weight markers with sizes given in kilodaltons (kDa); lanes b,e are control cells without vector; lanes c,f are uninduced cells; lanes d,g are cells following temperature induction for four hours (9,17). The major antigenic product of 32-kDa is arrowed.

observed as a cascade of protease activities involving both viral-encoded and cellular proteases (19). Such a mechanism may account for the processing of the IBDV precursor within the in vitro translation products (5) although the results do not distinguish whether the processing occurs by viral-encoded or cellular proteases. Similarly for the IBDV 32-kDa protein to appear as a stable product in E. coli (Figure 3) the precursor must be processed either by viral-encoded or bacterial proteases. Of the mature viral proteins the 37-kDa and 32-kDa appear as major structural proteins in the viral capsid whereas the 28-kDa protein is not efficiently packaged even though synthesized in equivalent

amounts (5). Thus a likely explanation for the in vitro translation and E. coli expression studies is that the 28-kDa protein may be a viral-encoded protease responsible for at least some events in precursor maturation. In this respect IBDV appears similar to the single stranded RNA genomes of picornaviruses such as foot and mouth disease virus (20,21) in which the single chain precursors are processed by an extensive series of proteolytic reactions, some of which are by viral-encoded proteases. There is some supporting evidence for such a processing mechanism in IBDV because larger proteins of 52-kDa and 41-kDa were identified as possible in vitro precursors to the native 37-kDa and 32-kDa structural proteins (5). Further evidence was gained using monoclonal antibodies on western blots of denatured viral proteins which identified a minor 55-kDa band immunogenically related to the 32-kDa structural protein (9,22). Although it seems likely that the 41-kDa protein is an extension of the 37-kDa antigen and the 55-kDa precursor may contain the 28 + 32-kDa proteins, the precise series of post-transcriptional and post translational modifications which may occur in vivo will only be resolved by more extensive structural analysis of the fragments involved. Further experiments are in progress to prove whether the viral genome represents the unspliced translation template in vivo.

Predicted Structure of the Native Viral Proteins

We have defined the N-terminus of the 37-kDa protein as residue 1 in Figure 2. The C-terminus of the 32-kDa protein must be within the eleven residues prior to termination codon and is most likely to be the unprocessed C-terminal Glu (residue 1011). From size estimates the N-terminus of the 32-kDa protein must be within residues 710-740. There are a few structural features in the protein sequence over these residues 710-740 in the predicted cleavage region between the 28 and 32-kDa protein and also in residues 441-542 as the predicted 37/28-kDa cleavage region. Dibasic residues are conveniently situated at residues 451-452 and 721-722 and cleavage at this site would excise a predicted 28.2-kDa protein. Dibasic residues are the frequent target for proteolysis of peptide hormone precursors (23) although a number of alternative cleavage specificities have been observed in viral-encoded proteases (24,25). The sequence A-X-A-A-S is repeated three times between residues 483-503 and also appears at 752-756. Although this region is highlighted in Figure 2 its potential as a cleavage site is

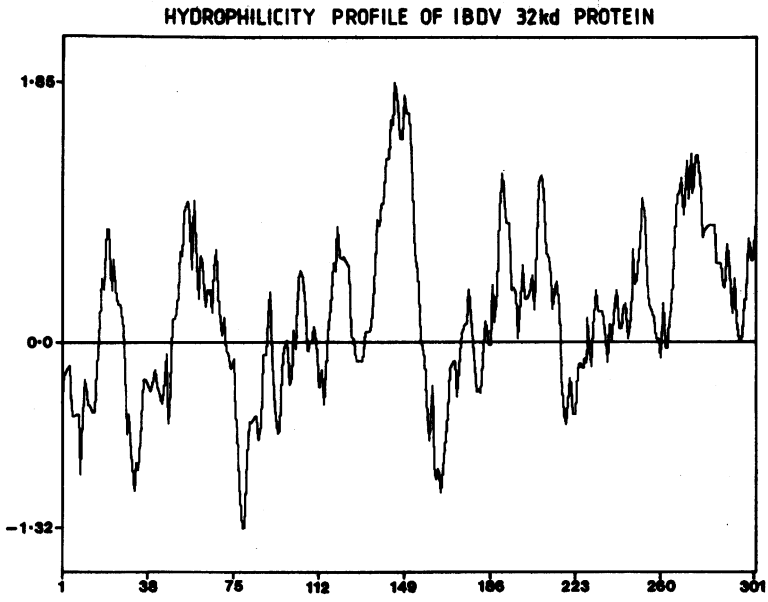

Figure 4
The hydrophilicity profile over the 32-kDa antigen encoded by the D6 clone. The sequence was numbered from residue 1 in the D6 clone which represents residue 702 in Fig. 2. Hydrophilicity calculations were smoothed by averaging over hexapeptide segments as described (31). Hence position 1 refers to the hexapeptide 702-707 in Fig. 2.

unknown and there is no homology of this structure to any cleavage sites in the NBRF protein database. It is apparent that further protein characterization on the small amounts extracted from the native virus or from translated products will be needed to resolve the definitive structures of the viral proteins and their precursors.

To establish whether any of the predicted viral proteins or the RNA template possessed homology to other structures the sequence data shown in Fig. 2 was searched against the entire NBRF protein database or the GENBANK nucleic acid database using an algorithm described by Wilbur and Lipman (26). Although no highly conserved regions were found in either database, residues 974-987 exhibited weak though significant homology to some DNA-binding proteins, notably histone H1.3 (residues 147-160, ref. 27), nonhistone chromosomal protein HMG-17 (residues 30-43, ref. 28), the capsid protein of Sindbis virus (residues 73-86, ref. 29) and the core antigen of Hepatitis B virus (residues 158-161,

ref. 30). These homologous sequences comprise a highly basic region with interspersed proline residues and are presumed to form a binding interaction with nucleic acids. Hence the basic C-terminal region of the 32-kDa protein may be involved in either packaging or stabilizing the RNA genome within the interior of the viral capsid.

Antigenicity of the 32-kDa protein

Since the earliest serum antibodies that appeared in chickens following infection with live virus or injection of an inactivated vaccine were specific for the 32-kDa polypeptide (7) it was proposed that the 32-kDa protein is a major immunogen of IBDV. In order to identify surface structures which are potential antigenic determinants the protein sequence predicted from the D6 clone was analysed by the hydrophilicity calculations of Hopp and Woods (31). In the analysis depicted in Figure 4 the hydrophilicity profile was smoothed by averaging over hexapeptide segments since this method has been used to successfully predict antigenic determinants exposed on the surface of a number of well-characterized immunogenic proteins (31). Peptides spanning residues 847-862 exhibit the highest peaks of hydrophilicity and are therefore likely to be surface determinants. This conclusion is supported by the observation that D1 also spans this region and the expressed fusion protein reacts strongly to monoclonal antibodies specific for the 32-kDa protein. The basic C-terminal region spanning residues 974-987 is also strongly hydrophilic although this basic sequence has been suggested to interact with the RNA genome in the interior of the viral capsid by homology to other DNA-binding proteins. From the data presented in Fig. 4 it is thus possible to begin targeting sequences for the development of a peptide based subunit vaccine (32).

Based on the sequence data, attempts are currently in progress to construct fusion proteins from which the IBDV protein can be excised specificically, and to construct novel vectors capable of expressing various regions of the IBDV genome with the aim of producing a host-protective subunit vaccine.

Acknowledgements

We thank Dean Hewish, Kevin Fahey, Peter McWaters and Con Sonza for providing monoclonal antibodies and Dehn Von Ahlefeldt, Paul Failla, Margaret Morgan, and Melissa Brown for expert technical assistance. This work was carried out under the Australian RDMC C1 containment conditions.

REFERENCES

1. Hirai, K. and Calnek, B.W. (1979). Infection and Immunity 25, 964-970.
2. Okoye, J.O.A. (1984). Vetinary Bulletin 54, 425-436.
3. Wyeth, P.J. and Cullen, G.A. (1979). Vet. Record 104, 188-193.
4. Lucio, B. and Hitchener, S.B. (1979). Avian Diseases 23, 466-478.
5. Azad, A.A., Barrett, S.A. And Fahey, K.J. (1985). Virology 143, 35-44.
6. Dobos, P., Hill, B.J., Hallett, R., Kells, D.T., Becht, H. And Teninges, D. (1979). J. Virol. 32, 593-605.
7. Fahey, K.J., O'Donnell, I.J. and Bagust, T.J. (1985). J. Gen. Virol. 66, 2693-2702.
8. Ruther, U. and Muller-Hill, B. (1983). EMBO. J. 2, 1791-1794.
9. Azad, A., Fahey, K., Barrett, S., Erny, K. and Hudson, P. (1986). Virology. 149, 190-198.
10. Maxam, A.M. and Gilbert, W. (1977). Proc. Natl. Acad. Sci. 74, 560-564.
11. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980). J. Molec. Biol. 143, 161-178.
12. Hudson, P., Haley, J., Cronk, M., Shine, J. and Niall, H. (1981).Nature 291, 127-131.
13. Staden, R. (1982). Nucleic Acids Res. 10, 4731-4751.
14. Goding, J.W. (1984). Recept. Biochem. Methodol. 3, 31-60.
15. McKern, N.M., O'Donnell, I.J., Stewart, D.J. and Clark, B.L. (1985). J. Gen. Microbiol. 131, 1-6.
16. Peterson, J.D., Nehrlich, S., Oyer, P.E. and Steiner, D.F. (1972). J. Biol. Chem. 247, 4866-4871.
17. Stanley, K.K. and Luzio, J.P. (1984). EMBO J. 3, 1429-1434.
18. Mertens, P.P.C. and Dobos, P. (1982). Nature 297, 243-246.
19. Tian, Y.C. and Shih, D.S. (1986). J. Virol. 57, 547-551.
20. Grubman, M.J. and Baxt, B. (1982). Virol. 116, 19-30.
21. Klump, W., Marquardt, O. and Hofschneider, P.H. (1984). Proc. Natl. Acad. Sci. 81, 3351-3355.
22. Fahey, K.J., O'Donnell, I.J. and Azad, A.A. (1985). J. Gen. Virol. 66, 1479-1488.
23. Steiner, D.F., Quinn, P.S., Chan, S.J., Marsh, J. and Tager, H.S. (1980). Ann. N.Y. Acad. Sci. 343, 1-16.
24. Henderson, L.E., Sowder,R., Copeland, T.D., Smythers, G. and Oroszlan, S. (1984). J. Virol. 52, 492-500.
25. Hanecak, R., Semler, B.L., Anderson, C.W. and Wimmer, E. (1982). Proc. Natl. Acad. Sci. 79, 3973-3977.
26. Wilbur, W.J. and Lipman, D.J. (1983). Proc. Natl. Acad. Sci. 80, 726-730.
27. Hsiang, M., Largman, C.R. and Cole, R.D. (1977) in 'The Molecular Biology of the Mammalian Genetic Apparatus', Vol. 1, pp.93-104, ed. P. Tso. Elsevier/North Holland.
28. Walker, J.M., Steam, C. and Johns, E.W. (1980). FEBS Lett. 112, 207-210.
29. Rice, C.M. and Strauss, J.H. (1981). Proc. Natl. Acad. Sci. 78, 2062-2066.
30. Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., Mackay, P., Leadbetter, G. and Murray, K. (1982). Nature 282, 575-579.
31. Hopp, T.P. and Woods, K.R. (1981). Proc. Natl. Acad. Sci. 78, 3824-3828.
32. Lerner, R.A. (1982). Nature 299, 592-596.