Structure, sequence and expression of the hepatitis delta (δ) viral genome

Kang-Sheng Wang*, Qui-Lim Choo*, Amy J. Weiner*, Jing-Hsiung Ou†, Richard C. Najarian*, Richard M. Thayer*, Guy T. Mullenbach*, Katherine J. Denniston‡, John L. Gerin‡ & Michael Houghton*§

* Chiron Corporation, 4560 Horton Street, Emeryville, California, 94608, USA
† Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA
‡ Division of Molecular Virology & Immunology, Georgetown University, 5640 Fishers Lane, Rockville, Maryland 20852, USA

Biochemical and electron microscopic data indicate that the human hepatitis δ viral agent contains a covalently closed circular and single-stranded RNA genome that has certain similarities with viroid-like agents from plants. The sequence of the viral genome (1,678 nucleotides) has been determined and an open reading frame within the complementary strand has been shown to encode an antigen that binds specifically to antisera from patients with chronic hepatitis δ viral infections.

THE delta (δ) hepatitis antigen (HDAg) was first recognized by Rizzetto and coworkers in 1977 as a novel nuclear antigen appearing in certain Italian carriers of the hepatitis B virus (HBV)[1]. Carriers expressing this HDAg exhibit a greater incidence of severe chronic active hepatitis and cirrhosis and the antigen is implicated in a substantial number of cases of fulminant hepatitis[2-4]. Chimpanzee transmission studies[5] indicate that a defective viral agent is associated with δ hepatitis and that this agent requires HBV or the woodchuck hepadna virus[6] in order to replicate. The hepatitis δ viral (HDV) agent replicates efficiently and suppresses helper replication which can lead to substantially higher titres of HDV relative to the hepadna virus[5,7]. HDV is now known to be endemic among the HBV carrier population in Italy[8] and has also been implicated in hepatitis cases throughout many parts of the world[9] where it occurs either as a result of a super-infection of the HBV carrier state or as an acute co-infection[8]. High risk groups include intravenous drug addicts and haemophiliacs who have received clotting factors derived from pooled blood donations[8].

In infectious sera, HDV particles of about 36 nm in diameter have been distinguished from the 42 nm Dane particle and 22 nm surface antigen moieties derived from HBV[10]. These particles appear to consist of an exterior HBV surface antigen and lipid coat which encapsidates HDAg and a δ-associated RNA molecule (HDV RNA)[11,12]. This RNA molecule has been estimated by gel electrophoresis to contain about 1,700 nucleotides and is proposed to be single stranded because of its ribonuclease sensitivity[13]. Recently, a small (164 base pairs) cDNA clone (pKD3) was obtained in one of our laboratories and shown to hybridize specifically with HDV RNA derived from a variety of infectious sera[14]. HDV RNA is presumed to be the viral genome because of its association with viral preparations and the correlation between infectivity titre and HDV RNA concentration in an acute phase chimpanzee serum[7,8]. We now report an extensive characterization of this HDV RNA.

Structure of RNA

Since it has been suggested that HDV RNA lacks a poly(rA) segment[13], we polyadenylated the RNA *in vitro* before synthesizing and cloning cDNA using the Okayama/Berg vector[15]. Recombinants were screened using an oligonucleotide hybridization probe derived from the sequence of pKD3. Two clones δ1 and δ2 were obtained containing 567 and 250 base pairs (bp) of cDNA sequence, respectively (excluding 'tails'). Strand-specific cDNA probes derived from clone δ1 were hybridized to the original HDV template RNA. Figure 1 demonstrates that only one of these probes hybridized to the HDV RNA thus indicating that this RNA is single-stranded. The hybridizing RNA species consists of a closely migrating RNA doublet of ~1,700 nucleotides in length.

To isolate additional clones, a cDNA library generated by random priming HDV RNA was screened with a hybridization probe derived from clone δ1. Positive clones were then used as hybridization probes to screen for new, overlapping clones in the cDNA library. In this way independent clones δ3b, δ4, δ7a,

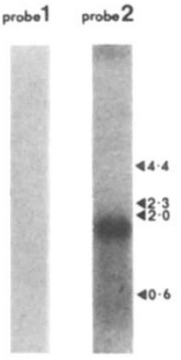


Fig. 1 Hybridization analysis of HDV RNA. Strand-specific hybridization probes 1 and 2 were derived from cDNA clone δ1 and hybridized individually to HDV RNA that had been previously electrophoresed through a denaturing RNA gel. The parallel migration of phage λ/*Hin*dIII restriction fragments (expressed in kbp) is indicated by arrows.

Methods. Strand-specific hybridization probes were prepared by cloning either a ~950 bp *Pvu*II/*Hin*dIII restriction fragment (probe 1) or a ~450 bp *Pvu*II/*Pst*1 fragment (probe 2) of clone δ1 into phage M13 vectors[39] to generate complementary single-stranded HDV templates. Hybridization probes were prepared by mixing 0.8 μg of each template DNA with 0.1 μg of hybridization probe primer (New England Biolabs) in 200 mM NaCl and incubating for 15 min at 37 °C after denaturing in a boiling water bath for 1 min. The annealed mixture was incubated at 15 °C for 2 h in 200 μl containing 50 mM Tris/Cl *p*H 7.5, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 50 μg/ml BSA, 0.1 mM dATP, dGTP and dTTP, 14 μM dCTP (1,000 C_i $mmol^{-1}$) and 250 units ml^{-1} Klenow *E. coli* DNA polymerase 1 prior to phenol/chloroform extraction and chromatography on a G50(M) Sephadex (Pharmacia) column (20 cm. × 0.7 cm diameter) run in 50 mM NaCl, 0.1% SDS. Probe eluting in the void volume was hybridized to a 'Northern blot'[40] of a 1% agarose gel containing 2.2 M formaldehyde[41] through which 10 ng HDV RNA and λ/*Hin*dIII restriction fragments had been electrophoresed. HDV RNA was derived from the serum of a chimpanzee during acute HDV infection and the preparation procedure included incubation with an excess of proteinase k and phenol/chloroform extraction[14,42,43]. The blot was pre-hybridized at 42 °C overnight in 5 × SSC (1 × SSC is 150 mM NaCl, 15 mM sodium citrate), 50% (v/v) formamide, 50 mM Na_2HPO_4 *p*H 6.5, 0.2% (w/v) SDS, 2 × Denhardt's solution (50 × is 1% (w/v) BSA, 1% (w/v) Ficoll-400, 1% (w/v) polyvinylpyrrolidone) and 250 μg ml^{-1} salmon testis DNA (sonicated and denatured) prior to hybridizing at 42 °C overnight in 5 × SSC, 50% (v/v) formamide, 20 mM Na_2HPO_4 *p*H 6.5, 0.1% SDS, 1 × Denhardt's, 10% (w/v) dextran sulphate, 100 μg ml^{-1} salmon testis DNA and 10^6 c.p.m. ml^{-1} of probe. Filters were washed at 65 °C in 0.1 × SSC, 0.1% SDS and autoradiographed at −80 °C with an intensifier screen.

§ To whom correspondence should be addressed.

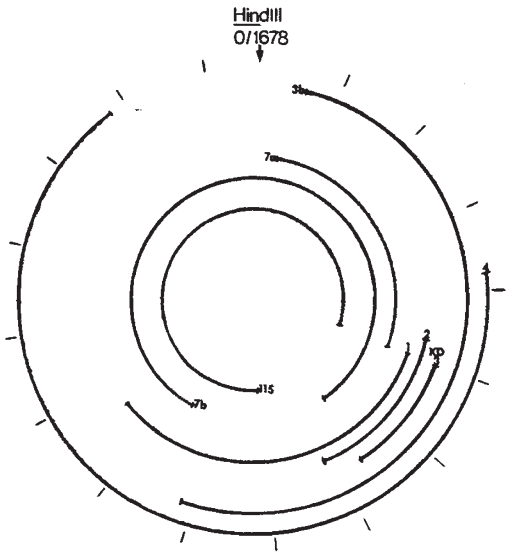


Fig. 2 Schematic alignment of the HDV cDNA sequences. The sequences of 7 independent HDV cDNA clones were determined (see Fig. 4) and aligned together with the initial clone pKD3. The scheme indicates the putative circular template RNA consisting of 1,678 nucleotides. The circle is shown as beginning in the centre of a unique *Hin*dIII restriction site. Increments of 100 nucleotides are indicated on the circumference.

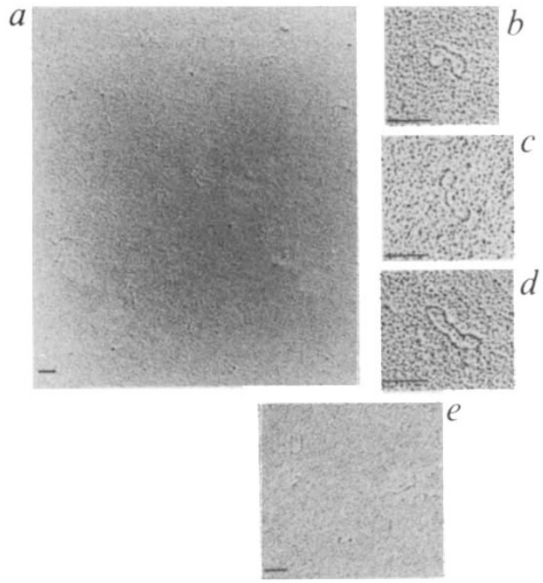


Fig. 3 Electron microscopy of HDV RNA. Panels *a–d*, completely denaturing conditions, *e*, weakly denaturing conditions. Several hundred molecules were observed under completely denaturing conditions. Circular and linear molecules were evident in similar proportions. Three RNA molecules of known length (777, 1,997 and 3,346 nucleotides) were used as standards under identical denaturing conditions and the length of these circles and linears was found to be 1,715 nucleotides (standard deviation of 175 from measurement of 22 different molecules) and 1,487 nucleotides (standard deviation of 290 from measurement of 31 different molecules) respectively. *a*, Typical field *b–d*, different, individual circular molecules. Several hundred molecules were observed under weakly denaturing conditions. All appeared as linear structures with an estimated size of 910 nucleotides (standard deviation of 70 from measurement of 38 different molecules). *e*, Typical field. Bar, 100 nm.

Methods. *a–d*, HDV RNA (50 ng) was heated at 70 °C for 15 min in 50 μl of 90% (v/v) formamide, 10 mM Tris/Cl, *p*H 8.2, 1 mM EDTA, then 1.5 μl of cytochrome *c* (1 mg ml^{-1}) was added. An aliquot (30 μl) was then immediately spread onto the hypophase of 60% (v/v) formamide, 1.5 mM Tris/Cl, *p*H 8.2, 0.15 mM EDTA, pre-warmed to 70 °C. 30 s after spreading (at which time the temperature was >50 °C), the surface film formed by the sample was picked up on a carbon-coated grid. The grid was stained (45 s) in 0.002% (w/v) uranyl acetate in ethanol before shadowing with platinum and viewing in a Phillips 400 electron microscope[16]. *e*, HDV RNA (50 ng) was suspended at room temperature in 50 μl of 40% (v/v) formamide, 10 mM Tris/Cl, *p*H 8.2, 1 mM EDTA and 60 μg ml^{-1} cytochrome *c* and spread onto the hypophase of 10% (v/v) formamide, 1 mM Tris/Cl *p*H 8.2, 0.1 mM EDTA and viewed in a Phillips 400 electron microscope[44].

δ7b and δ115 were identified and shown to contain cDNA inserts of 829, 1,123, 474, 1,378 and 1,362 bp, respectively (excluding G-C 'tails').

Surprisingly, comparison of the nucleotide sequences of the cloned cDNAs suggested that the template RNA for at least some of them was a circular molecule of 1,678 nucleotides (Fig. 2). A linear model would require a molecule comprising ~2,180 nucleotides and containing terminal direct repeats of ~510 nucleotides. However, the HDV RNA migrates faster than 18S ribosomal RNA (1,869 nucleotides) on denaturing formaldehyde gels (ref. 14 and K.S.W., unpublished data) indicating a size of around 1,700 nucleotides. A hypothesis involving heterogeneous, permuted linear HDV RNA molecules could be consistent with the cloning and sequence data. A significant level of base substitution was observed in the sequences of overlapping regions of different clones but no insertion or deletion events were found, suggesting that the templates used to generate these seven clones are of uniform size.

The batch of HDV RNA used for cDNA cloning was examined in the electron microscope under completely denaturing conditions[16]. Circular and linear molecules were present in approximately similar proportions (see Fig. 3, *a–d*). The mean sizes of circles and linears observed in the electron microscope under these conditions were calculated to be 1,715 and 1,487 nucleotides respectively, values close to that of 1,678 nucleotides from sequence analysis and ~1,700 nucleotides from gel electrophoresis. The estimated smaller size of the linear molecules seen in the electron microscope probably reflects the presence of some degraded molecules, whereas circles were presumably all of full length.

Hybridization analysis of HDV RNA using probes derived from the random-primed cDNA clones yielded results similar to those shown in Fig. 1 (K.S.W., unpublished data) indicating that HDV RNA is completely single-stranded. It seems likely therefore that HDV contains a single-stranded, circular RNA molecule of 1,678 nucleotides, presumably covalently closed since cDNA synthesis appears to proceed around the molecule (Fig. 2) and circular forms are visible in the electron microscope under completely denaturing conditions for RNA (Fig. 3, *a–d*).

When electron microscopy of this RNA was conducted under much less denaturing conditions only linear molecules were seen (Fig. 3, panel *e*), of mean size ~910 nucleotides.

Nucleotide sequence

Figure 4 shows the nucleotide sequence of HDV RNA as deduced from analysis of the cDNAs in clones δ1, δ2, δ3b, δ4, δ7a, δ7b and δ115. The region corresponding to the short cDNA sequence in clone pKD3 was derived from three independent clones generated in this study and differs in several positions from the sequence of pKD3[14].

A prominent feature of the HDV RNA sequence is the high G+C content (60%) which results in the formation of stable secondary structure. Computer analysis of the sequence indicates an inherent ability of HDV RNA to self-anneal. Figure 5*a* shows a computer line matrix analysis where the entire HDV RNA sequence is scanned for homologies of 6 nucleotides or

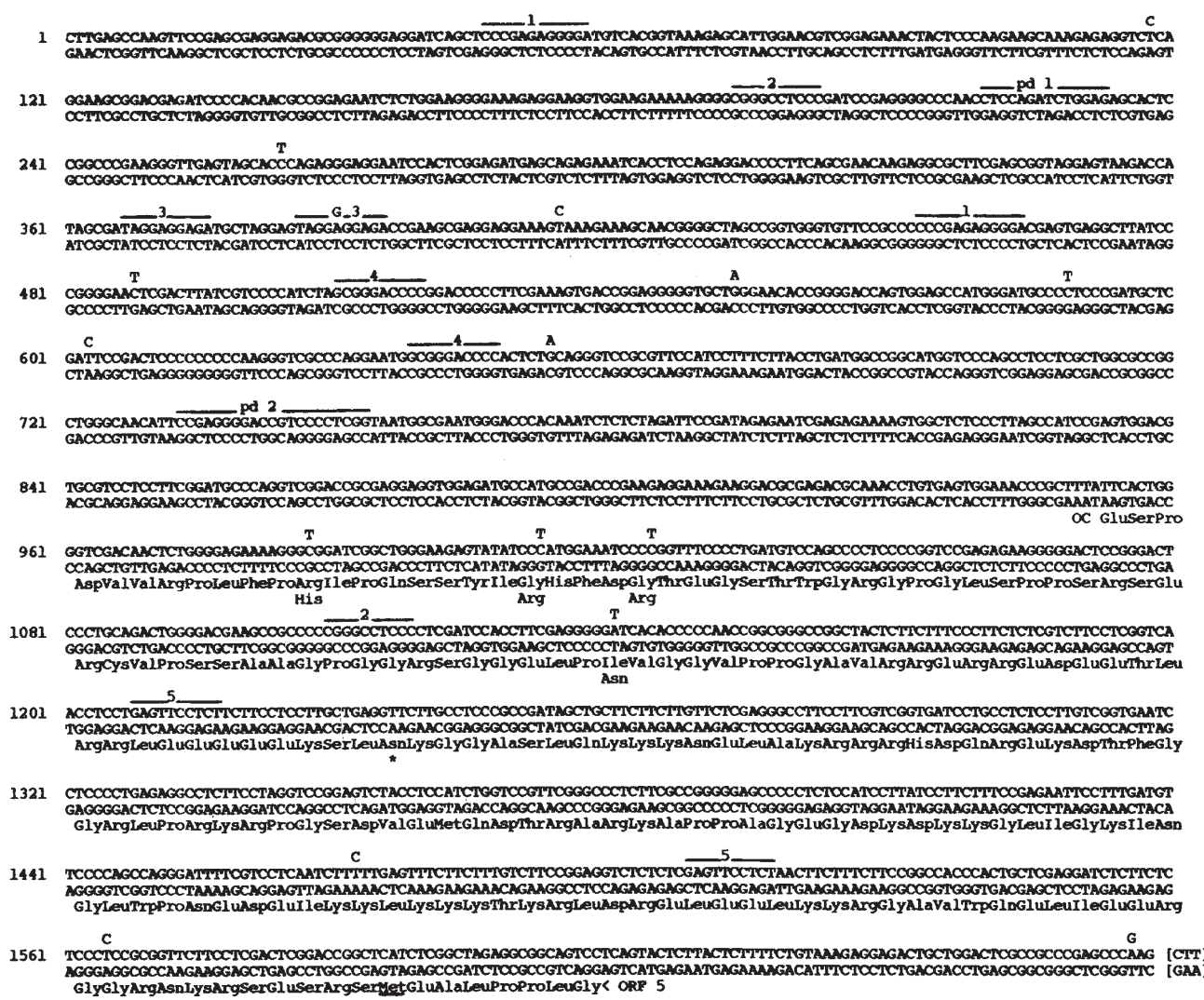


Fig. 4 Nucleotide sequence of HDV RNA. The sequence was derived from 7 different cDNA clones and is shown in double-stranded DNA form. The upper strand corresponds to the sequence of the single-stranded HDV RNA molecule (T residues represent the U residues in the HDV RNA). The sequence is shown beginning and ending at the centre of a unique *Hin*dIII site. Base substitutions observed between clones are indicated, as are direct repeats (1-5) of at least 10 nucleotides and two long palindrome-like sequences (pd 1, 2). The translation of ORF 5 in the lower anti-HDV RNA strand is shown . A potential site for *N*-linked glycosylation is indicated by an asterisk. An ORF in the upper HDV RNA sequence (nucleotides 1,477-1,678 and 1,678-2) exhibits the greatest region of nucleotide homology with the HBV (ayw) genome. The amino acids encoded by nucleotides 1,483-1,557 of this ORF display the most homology with a carboxy terminal region of the HBV polymerase Ser Leu Ser Ser Gly Gly Leu Ser Arg Val Pro Leu Thr Ser Phe Phe Arg Pro Pro Thr Ala Arg Gly Ser Leu (underlined residues indicate HDV ORF homologies with residues 780-804 of the HBV polymerase[18,19]).

Methods. Clones δ1 and δ2: the RNA (150 ng) was polyadenylated by pre-heating at 70 °C for 2 min in 50 mM Tris/Cl *p*H 7.9, 2 mM EDTA and incubating at 37 °C for 2 min in 50 mM Tris/Cl *p*H 7.9, 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 250 mM NaCl, 40 μM rATP, 2 mM DTT, 1,000 units ml^{-1} RNAase inhibitor (Promega) and 100 units ml^{-1} *Escherichia coli* poly(rA) polymerase (BRL) (50 μl). The RNA was extracted with phenol/chloroform and precipitated overnight at −20 °C with 2.5 volumes of absolute ethanol, recovered by centrifugation and resuspended in H_2O, then cloned into the Okayama/Berg vector[15] using *E. coli* MC1061[45] as host. Ampicillin-resistant transformants were colony-hybridized[46] using an oligonucleotide (5′ GAGTCGGAATCGAGCATCGGGAAGGGCATC) as hybridization probe (derived from the sequence of pKD3[14]). The probe was labelled by incubating 20 pmol at 37 °C for 1 h in 40 μl of 0.1 mM EDTA, 50 mM Tris Cl *p*H 9.0, 10 mM $MgCl_2$, 0.1 mM spermidine, 1 mM DTT containing 200 μCi [^{32}P]ATP(>5,000 Ci $mmol^{-1}$) and 700 units ml^{-1} T4 polynucleotide kinase. After extraction with phenol and chloroform, the aqueous phase was bound to a Sep-Pak C18 cartridge (Millipore), eluted with 50% (v/v) CH_3OH, 50 mM CH_3COONH_4, *p*H 7.5, dried and resuspended in H_2O. The probe was hybridized at ~10^6 c.p.m. ml^{-1} using the hybridization and washing conditions described in Fig. 1. Positive clones δ1 and δ2 were thus identified. HDV clones 3b, 4, 7a, 7b and 115: HDV RNA (450 ng) was denatured in 10 mM CH_3HgOH at room temperature for 10 min and single strand cDNA was synthesized in a 50 μl reaction containing 50 mM Tris Cl *p*H 8.3, 10 mM $MgCl_2$, 70 mM KCl, 0.5 mM dATP, dGTP and TTP, 0.2 mM [^{32}P] dCTP (15 Ci $mmol^{-1}$), 600 units ml^{-1} RNAase inhibitor (Promega), 14 mM β-mercaptoethanol, 5 mM DTT, 100 units ml^{-1} AMV reverse transcriptase (Life Sciences), 50 μg ml^{-1} actinomycin D, 0.01% (v/v) Triton X-100 and 4 μg ml^{-1} random primers derived from a limit DNAase 1 digest of calf thymus DNA[47]. After incubation at 37 °C for 1.5 h, the cDNA/RNA hybrids were extracted with phenol/chloroform, 5 μg chimpanzee liver RNA was added as carrier and the hybrids were precipitated with ethanol. Subsequent cloning steps have been described in detail elsewhere[48]. Briefly, RNA was degraded (0.3 N NaOH, 1 h, 50 °C) and double stranded cDNA was prepared by allowing self-priming with reverse transcriptase. Hairpins were cleaved with S_1 nuclease and oligo(dC) tails were added using terminal transferase and annealed with G-tailed, *Pst*1-cut pBR322 (New England Nuclear) and transformed into *E. coli* MC1061[45]. Ampicillin-sensitive, tetracycline-resistant recombinants were colony-hybridized using a 'nick translated' 435 bp *Nco*I cDNA restriction fragment of clone δ1 as hybridization probe. The cDNA inserts of positive clones (δ4 and δ115) were used as hybridization probes for the same library, giving clones 3b, 4, 7a, 7b and 115. Following restriction mapping, various restriction fragments of the cDNA inserts from each clone were sub-cloned into M13 vectors (mp18 and mp19)[39] and sequenced using the dideoxy chain termination method[49]. The sequence shown was deduced from complementary strands of at least two overlapping clones. Areas of uncertainty were confirmed using the techniques of Maxam and Gilbert[50].

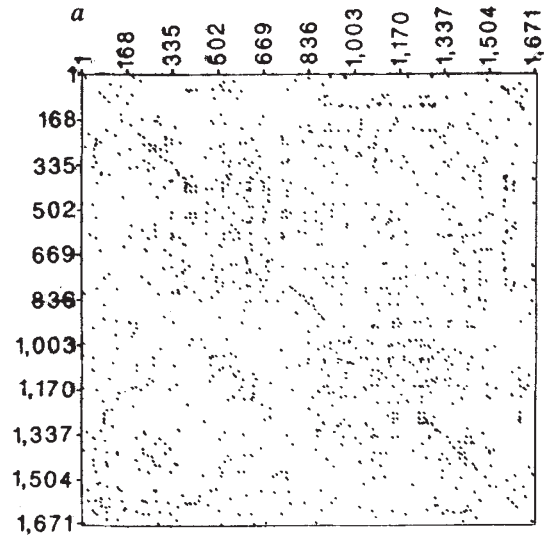
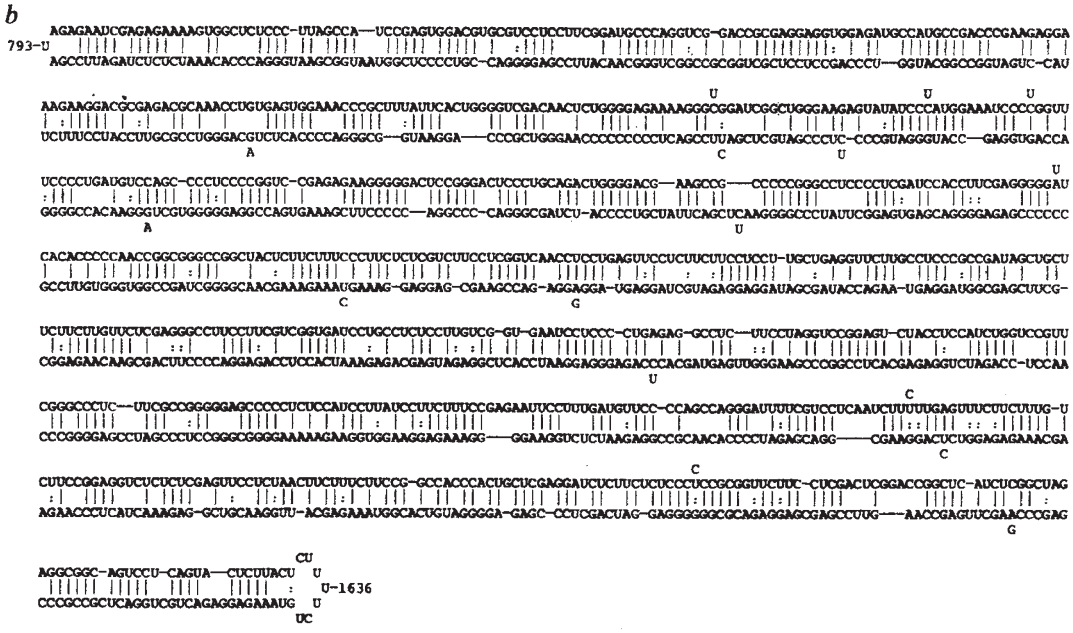


Fig. 5 Self-complementarity of HDV RNA. *a*, Computer line matrix, *b*, Putative covalently-closed, double-stranded HDV RNA structure. The matrix was obtained by scanning for homologies (≥6 nucleotides) between HDV RNA and its complementary sequence. The vertical axis (1→1,678) represents the sequence of the entire HDV RNA molecule (the upper strand in Fig. 4 from nucleotides 793–1,678 followed by 1–792) while the horizontal axis (1→1,678) represents the entire complementary sequence of the HDV RNA (the lower strand in Fig. 4 from nucleotides 792 back to 1 followed by nucleotides 1,678 back to 793). *b*, The complementing halves of the HDV RNA sequence indicated in Fig. 5*a* were arranged by computer for maximum base-pairing. Vertical lines, base-pairs; colons, potential G:U partial matches[17]. Base substitutions between different cDNA clones are indicated. Nucleotides are numbered according to the scheme used in Fig. 4.



more with its complementary sequence. Homology is indicated by a short line of length depending on the size of the homologous region. This analysis indicates regions of intra-molecular and self-complementarity, and putative self-annealing within the HDV RNA molecule itself. Such regions extend throughout the entire HDV RNA sequence as can be seen from the semi-continuous diagonal line that begins at the origin and extends throughout the matrix figure. Figure 5*a* indicates that nucleotides 793–1,632 (Fig. 4) of the HDV RNA molecule (*y* coordinates 1–839 of Fig. 5*a*) can extensively complement the other half of the HDV RNA sequence (*x* coordinates 1–839 of Fig. 5*a*). Using computer analysis to obtain maximum complementation in this region a highly base-paired, covalently-closed structural model for the HDV RNA molecule can be proposed (Fig. 5*b*). Sixty-one per cent of the nucleotides are base-paired in this putative rod-like structure and a further 6% can take part in G-U matches[17]. The linear molecules observed in the electron microscope under weak denaturing conditions (Fig. 3*e*) were approximately half the size of the circular molecules seen under completely denaturing conditions, consistent with their being double-stranded, rod-like structures.

The HDV RNA sequence and the complementary anti-HDV RNA strand contain numerous open reading frames (ORFs) the largest of which are indicated in Table 1. Many of these ORFs contain potential initiator ATG codons which, if responsible for translation initiation, could encode polypeptides of up to 215 amino acids before taking into account possibilities of splicing and/or frameshifting events.

In the comparison of the HDV RNA sequence with all other known viral and viral-like sequences (using computer analysis), the most significant homology detected was with the S strand of the ayw strain of HBV[18]. Allowing for many gaps in the alignment, a low level of homology was apparent which had some statistical significance ($P = 0.02$) compared with the homologies obtained using randomized nucleotide sequences (data not shown). The greatest region of homology is in a small ORF in HDV RNA (nucleotides 1,477–1,678/1,678–2; Fig. 4) with a coding potential of 90 amino acids, although none of these is methionine or hence a possible translation initiation site. However, part of the potential polypeptide encoded by this ORF has some homology with the carboxy terminus of the HBV (ayw) polymerase[18,19] with the 25 amino acids encoded by

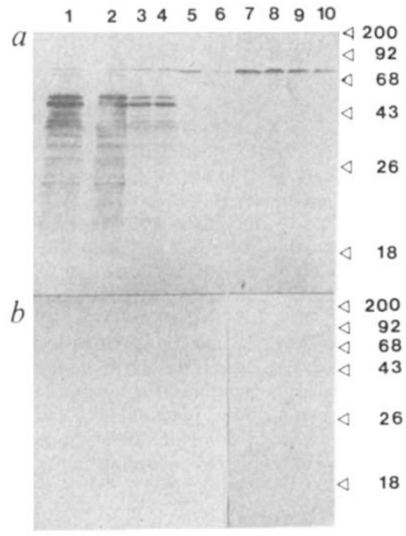


Fig. 6 Immunological analysis of polypeptides expressed in HDV cDNA recombinant bacteria. ORFs 5 and 6 were expressed in the form of fusion polypeptides with the human enzyme superoxide dismutase (SOD) under the control of the inducible 'Tac' promoter[20], then reacted with either *a*, antiserum from a patient chronically infected with HDV or *b*, a pool of control antisera from previous infections with either hepatitis A, B or Non-A, Non-B viral agents. The bacteria were transformed with: lanes 1 and 2, pSOD-ORF 5 (induced with IPTG); lanes 3 and 4, pSOD-ORF 5 (no IPTG); lane 5, parent pSOD vector (induced with IPTG); lane 6, parent pSOD vector (no IPTG); lanes 7 and 8, pSOD-ORF 6 (induced with IPTG); lanes 9 and 10, pSOD-ORF 6 (no IPTG). The migration of protein standard (expressed as r.m.m. $\times 10^{-3}$) is indicated.

Methods. HDV cDNA restriction fragments corresponding to the regions of ORF 5 and ORF 6 were inserted into the region of plasmid pSOD16cf2[20] that specifies the carboxy terminus of SOD, allowing the synthesis of fusion proteins. To make the ORF 5 fusion protein, pSOD16cf2 was cleaved with *Nco*I, 'blunt-ended' with the Klenow fragment of *E. coli* DNA polymerase I and digested with *Sal*I prior to extracting the large *Nco*I/*Sal*I vector fragment from agarose gels following electrophoresis. Clone δ115 was digested with *Sst*II, 'blunt-ended' and restricted with *Sal*I; the fragment (~600 bp) was recovered from agarose gels. The frgments were ligated using T4 DNA ligase and transformed into *E. coli* strain D1210[51]. Recombinants were identified by restriction mapping and sequence determination of the junction between the human SOD cDNA and the HDV cDNA. Recombinant bacteria expressing ORF 6 fusion proteins were made in a similar way using a *Sma*I/*Eco*RI digested vector fragment and a 622 bp *Sma*I/*Eco*RI fragment derived from clone δ4. Overnight cultures of recombinant bacteria were diluted 100× into L-broth containing 100 μg ml^{-1} ampicillin and grown at 37 °C to an OD_{650} of 0.6 before either extraction or addition of 1 mM IPTG (Sigma) for 4 h to maximize expression before extraction. Bacteria were pelleted, resuspended in 10% (v/v) glycerol, 50 mM DTT, 3% (w/v) SDS, 60 mM Tris/Cl *p*H 6.8 and freeze thawed (3 cycles). After incubation in a boiling water bath for 5 min, samples were electrophoresed through 12% (w/v) polyacrylamide gels containing SDS[52] and electroblotted onto nitrocellulose filters[53]. The filters were pre-incubated with 5% (v/v) goat serum, then incubated with a 1:300 dilution of antisera in phosphate-buffered saline (PBS) containing 0.3% Tween 20 (Sigma) and 5% (v/v) goat serum for 1 h at room temperature. The filters were then incubated with a 1:200 dilution of horse radish peroxidase conjugated to goat anti-human IgG in the presence of the chromogen 4-chloro-1-naphthol (Biorad).

nucleotides 1,483–1,557 showing 52% homology with a carboxy-terminal region of HBV polymerase (Fig. 4, legend). Although there are a number of direct repeats and palindrome-like sequences in HDV RNA (Fig. 4) of which some may be important in replication, the direct repeats DR1 and DR2[19] that play an important role in replication of hepadna viruses are not present.

Expression of HDV

In an attempt to identify the ORF(s) coding for HDAg (previously termed δ antigen), two long ORFs (ORFs 5 and 6; Fig. 4 and Table 1) apparent in the sequence complementary to HDV RNA were selected for study. These ORFs were sub-cloned into a bacterial plasmid which uses the strong inducible 'Tac' promoter to express the ORFs as a fusion protein containing the human superoxide dismutase polypeptide at the amino terminus and the HDV-encoded polypeptides at the carboxy terminus[20]. Fusion proteins synthesised in this way were transferred to nitrocellulose after gel electrophoresis and then monitored for specific binding to an antiserum derived from a patient with a chronic HDV infection. Figure 6*a* shows that ORF 5 but not ORF 6 fusion polypeptides specifically bound HDV antiserum. A different batch of human HDV antiserum also binds the ORF 5 fusion proteins (Q.L.C., A.J.W., unpublished observation) whereas control antisera did not (Fig. 6*b*). Both HDV antisera have been shown to contain antibodies capable of binding to HDV viral polypeptides[21].

The estimated size of the largest major immunoreactive ORF 5 polypeptide observed in Fig. 6*a* (49,000 daltons) is consistent with a fusion polypeptide containing 154 amino acids of superoxide dismutase and 205 amino acids specified by ORF 5. Translation of ORF 5 in the other two reading frames (or in the opposite orientation) would produce much smaller polypeptides because of the presence of stop codons. The size heterogeneity observed with the ORF 5 fusion products may be a result of bacterial proteolysis. The failure to detect specific immunoreactive ORF 6 fusion polypeptides was not due to poor bacterial expression since, when monitored for binding to rabbit antiserum raised against human superoxide dismutase, both ORF 5 and ORF 6 fusion polypeptides were present at similar levels (A.J.W., unpublished data). Thus ORF 5 contributes to the synthesis of an immunogenic HDV polypeptide, which has unusual structural features. Clusters of basic and acidic residues appear throughout most of the molecule, except for the carboxy terminus which is much less charged and more hydrophobic in character. With an overall positive charge, the molecule could bind nucleic acid, as expected for the HDAg antigen, which is found inside the viral particle with the viral RNA. The hydrophobic carboxy terminus may interact with the coat of the HDV particle, which consists of HBV surface antigens and host-derived lipid components[11,12]. There is also a potential *N*-linked glycosylation site[22] in the putative ORF 5 polypeptide (Fig. 4).

Conclusion

Our evidence suggests that HDV contains a single-stranded covalently-closed circular RNA molecule of 1678 nucleotides which can form extensive secondary structure. The observation that HDV cDNA encodes an antigen that specifically binds antisera from patients with chronic HDV infections confirms previous data suggesting the genomic character of HDV RNA.

The unusual properties of the HDV genome resemble those of a number of different viroid-like molecules that cause a variety of diseases in plants. Viroids[23,17] are small single-stranded circular RNA molecules of high G-C content that form double-stranded rod-like structures[24] which are not encapsidated and do not require a helper agent for replication. Virusoids, a class of viroid-like molecules, are one of the two essential components of the bipartite RNA genomes of recently discovered plant viruses[25]. The other component is a large linear RNA molecule that may encode the replication function for both molecules. Finally, the satellite of the Lucerne Transient Streak virus[26], previously considered a tomato virusoid[27], has a viroid-like structure and, as noted earlier[28], is similar to HDV in requiring a largely unrelated helper virus (which it inhibits) for replication.

The sequence GAAAC occurs in all of these viroid-like molecules (except the hop stunt viroid[29]) and in the HDV genome (nucleotides 88–92, Fig. 4). The sequence GAUUUU also exists in the genomes of virusoids and HDV (nucleotides 1453–1458, Fig. 4). In addition, the HDV genome displays some homology (nucleotide regions 75–94 and 552–571 (Fig. 4) both

Table 1 Large open reading frames (ORFs) within HDV RNA and the complementary anti-HDV RNA strand

Strand	ORF number	Translational frame	Nucleotide* position	Total no. of amino acids	No. of amino acids beginning with first methionine
HDV RNA	1	2	539	165	152
HDV RNA	2	3	786	169	134
HDV RNA	3	2	1,607	120	0
HDV RNA	4	3	1,296	115	68
Anti-HDV RNA	5	1	1,618	222	215
Anti-HDV RNA	6	2	1,140	221	179/34†
Anti-HDV RNA	7	3	506	148/80	148/80‡
Anti-HDV RNA	8	3	1,340	109	0
Anti-HDV RNA	9	1	91	101	0

Only ORFs longer than 300 nucleotides are shown.
* The position of the first nucleotide in each open reading frame is indicated according to the numbering shown in Fig. 4.
† Ambiguity arising from clonal heterogeneity at position 1012 (Fig. 4).
‡ Ambiguity arising from clonal heterogeneity at position 264 (Fig. 4).

display 60–65% homology) with a viroid central conserved region, possibly involved in the processing of replication intermediates[30]. The functional significance of these homologies and of the general similarities with plant agents remains to be established. It is possible that HDV replicates in a 'rolling-circle' mechanism through the action of a replicase activity, similar to that suggested for viroids[31].

Although the sequences of the numerous HDV cDNA clones give no indication of size heterogeneity in the original HDV template RNA, two closely-migrating RNA species were observed following gel electrophoresis under denaturing conditions (Fig. 1). The species are highly homologous as hybridization probes remained annealed to the doublet after very stringent washing (Fig. 1). It is possible that as has been shown for plant viroids[16,29,32], the two species represent circular and 'nicked' linear forms of the same molecule. This hypothesis is supported by the observation of circles and linear molecules of similar size in the electron microscope under completely denaturing conditions (Fig. 3*a*) and the generation of clones $\delta 1$ and $\delta 2$ following an initial polyadenylation of HDV RNA termini.

It should also be noted that in the electron microscope a few circular molecules appear to possess a tail (Fig. 3*d*). Although this is probably an artefact of the cytochrome C background, the possibility that a subset of HDV RNA molecules have circular lariat structures similar to those previously described for certain processed introns[33,34] cannot be excluded.

A low level of homology is apparent between the HDV genome and the S strand of HBV (strain ayw). The implications of this are unclear as yet in the absence of further information about the replication of HDV. The degree of homology is consistent with our recent work showing that the genomes of HDV and HBV do not cross-hybridize[13] even under low stringency conditions[35].

There are many ORFs in both the genomic and anti-genomic strands of HDV. Interestingly, an ORF in the anti-genomic strand (ORF 5; Fig. 4 and Table 1) appears to encode an immunogenic viral antigen that is specifically recognized by antisera derived from patients with chronic HDV infections, whereas another ORF in the same anti-genomic strand (ORF 6, Table 1) does not. A similar negative result has also been obtained with an ORF (ORF 1, Table 1) in the genomic strand (A.J.W., unpublished data). Combined with the fact that eukaryote ribosomes cannot translate circular RNA molecules[36,37] and the observed inability of an *in vitro* rabbit reticulocyte lysate to translate the HDV genome (K.S.W., unpublished data), our data suggest that HDV is a negative-stranded virus. However, further studies are necessary before we can rule out any coding function for the genomic strand.

If translation of ORF 5 initiates at the first methionine codon and terminates at the first in-frame stop codon a polypeptide of 215 amino acids (about 25,000 daltons) would be synthesized (Fig. 4). Although earlier work using gel filtration analysis indicated that the molecular weight of HDAg was about 68,000[38], very recent analyses employing denaturing electrophoresis and Western blot techniques indicate that the HDV particle contains antigens of around 24,000 and 27,000 daltons[21] that bind specifically to the antisera used in this study. This data is therefore consistent with ORF 5 in the anti-genomic strand being responsible for the synthesis of at least one of these HDV polypeptides.

We thank S. Brown-Shimer for sequencing work, G. Bell, P. Luciw, R. Sanchez-Pescador, P. Valenzuela and M. Power for critical evaluation of the manuscript and assistance with computer analysis, B. Baroudy for helpful discussion, Toni H. Jones for typing the manuscript, and L. Overby for advice and encouragement. This work was funded by Chiron Research Laboratories, Emeryville, California.

Received 6 June; accepted 20 August 1986.

1. Rizzetto, M. *et al. Gut* **18**, 997–1003 (1977).
2. Jacobson, I. M. *et al. Hepatology* **5**, 188–191 (1985).
3. Govindarajan, S., Chin, K. P., Redeker, A. G. & Peters, R. L. *Gastroenterology* **86**, 1417–1420 (1984).
4. Arico, S. *et al. Lancet* **ii**, 356–358 (1985).
5. Rizzetto, M. *et al. J. infect. Dis.* **141**, 590–602 (1980).
6. Ponzetto, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2208–2212 (1984).
7. Ponzetto, A. *et al. J. infect. Dis.* (in the press).
8. Rizzetto, M. *Hepatology* **3**, 729–737 (1983).
9. Purcell, R. H. & Gerin, J. L. in *Viral Hepatitis and Delta Infection* (ed. Verme, G., Bonino, F. & Rizzetto, M.) 113–119 (Liss, New York, 1983).
10. Rizzetto, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 6124–6128 (1980).
11. Bonino, F. *et al. Hepatology* **1**, 127–131 (1981).
12. Bonino, F. *et al. Infect. Immun.* **43**, 1000–1005 (1984).
13. Hoyer, B. *et al.* in *Viral Hepatitis and Delta Infection* (ed. Verme, G., Bonino, F. & Rizzetto, M.) 91–97 (Liss, New York, 1983).
14. Denniston, K. J. *et al. Science* **232**, 873–875 (1986).
15. Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161–170 (1982).
16. Randles, J. W. & Hatta, T. *Virology* **96**, 47–53 (1979).
17. Haseloff, J., Mohamed, N. A. & Symons, R. H. *Nature* **299**, 316–321 (1982).
18. Galibert, F., Mandant, E., Fitoussi, F., Tiollais, P. & Charnay, P. *Nature* **281**, 646–650 (1979).
19. Tiollais, P., Pourcel, C. & Dejean, A. *Nature* **317**, 489–495 (1985).
20. Steimer, K. S. *et al. J. Virol.* **58**, 9–16 (1986).
21. Bergmann, K. F. & Gerin, J. L. *J. infect. Dis.* (in the press).
22. Walter, P., Gilmore, R. & Blobel, G. *Cell* **38**, 5–8 (1984).
23. Diener, T. O. *Adv. Virus Res.* **28**, 241–283 (1983).
24. Riesner, D. *et al. Biophys. struct. Mech.* **9**, 145–170 (1983).
25. Haseloff, J. & Symons, R. H. *Nucleic Acids Res.* **10**, 3681–3691 (1982).
26. Jones, A. T., Mayo, M. A. & Duncan, G. H. *J. gen. Virol.* **64**, 1167–1173 (1983).
27. Keese, P., Bruening, G. & Symons, R. H. *FEBS Lett.* **159**, 185–190 (1983).
28. Kaper, J. M. & Tousignant, M. E. *Endeavour, New Ser.* **8**, 194–200 (1984).
29. Ohno, T., Takamatsu, N., Meshi, T. & Okada, Y. *Nucleic Acids Res.* **11**, 6185 6197 (1983).
30. Diener, T. O. *Proc. natn. Acad. Sci. U.S.A.* **83**, 58–62 (1986).
31. Branch, A. D. & Robertson, H. D. *Science* **223**, 450–455 (1984).
32. Hashimoto, J., Suzuki, K. & Uchida, T. *J. gen. Virol.* **66**, 1545–1551 (1985).
33. Halbreich, A., Pajot, P., Foucher, M., Grandchamp, C. & Slonimski, P. *Cell* **19**, 321–329 (1980).
34. Zeitlin, S. & Efstratiadis, A. *Cell* **39**, 589–602 (1984).

35. Weiner, A. J. *et al. J. med. Virol.* (in the press).
36. Kozak, M. *Nature* **280**, 82-85 (1979).
37. Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J *Eur. J. Biochem.* **114**, 221-227 (1981).
38. Rizzetto, M., Shih, J. W-K. & Gerin, J. L. *J. Immun.* **125**, 318-324 (1980).
39. Messing, J. *Meth. Enzym.* **101**, 20-37 (1983).
40. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
41. Goldberg, D. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5794-5798 (1980).
42. Ponzetto, A. *et al. Hepatology* (in the press).
43. Smedile, A. *et al. Hepatology* (submitted).
44. Davis, R. W., Simon, M. & Davidson, N. *Meth. Enzym.* 413-428 (1971).
45. Casabadan, M. & Cohen, S. *J. Molec. Biol.* **138**, 179-207 (1980).
46. Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961-3965 (1975).
47. Taylor, J. M., Illmensee, R. & Summers, J. *Biochim. biophys. Acta* **442**, 324-330 (1976).
48. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning* 229-242 (*Cold Spring Harbor Laboratory* (New York, 1982).
49. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
50. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
51. Sadler, J. R. *et al. Gene* **8**, 279-300 (1980).
52. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
53. Towbin, H. T., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).

LETTERS TO NATURE

Double clusters and gravitational lenses

C. S. Crawford, A. C. Fabian & M. J. Rees

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

Clusters of galaxies can act as gravitational lenses[1] which produce double quasars with splittings as large as 1–2 arc min. The potential well of the cluster must be deep enough on the scale of the image to focus the light, which means that the velocity dispersion within the cluster must be high and the core radius small. These requirements are reduced if there are two (or more) clusters along the line of sight. Paczynski and Gorski[2] have modelled the triple image produced by two clusters without pursuing the probability of such an alignment in any detail. Here we predict the number of double quasar images in the sky that are produced by double clusters. The clustering of clusters[3] and the occurrence of binary clusters[4,5] enhances the probability of double clusters well above that expected from a purely random spatial distribution. Several double images are expected. Double clusters should be considered for apparent double quasar images such as Hazard 1146+111 B,C (refs 6, 7) before cosmic strings[8] or supermassive black holes[9] are required. Although H1146+111B and C may not be due to a gravitational lens[10,11], the discovery of a *bona fide* wide double quasar image will be important in the study of the evolution of the clustering of clusters.

We have used two approximate methods to estimate the number of wide double images. The first works from the number density of rich clusters[12] and the clustering of clusters[3]. There is no strong richness/velocity dispersion relation[13,14] but we shall assume that two clusters of Abell richness class 2 contribute an effective line-of-sight velocity dispersion, $\sigma_{\parallel}$, of 2,000 km s^{-1}. The angular separation of the gravitational images is $\sim 2\,(\sigma_{\parallel}/2{,}000\ \mathrm{km\ s^{-1}})^2$ arc min. This requires that they each have $\delta_{\parallel}$ of $\sim$1,400 km s^{-1}. The present number density of clusters of this richness and greater[12] is, $n(0) \approx 6\times10^{-8}$ Mpc^{-3} (we adopt $H_0 = 50$ km s^{-1} Mpc^{-1}). A similar result is obtained using the Piccinotti *et al.*[15] X-ray sample and the Quintana-Melnick[16] X-ray luminosity/velocity dispersion correlation. The fraction of the sky covering by double clusters aligned within an angle closer than the observed image separation θ (assuming no spatial correlation) is

$$f = \left(\frac{4c^3}{H^3}\pi\theta^2 n(0)\int_{\zeta_1}^{\zeta_2}(\zeta^{-3/2}+\zeta^{-5/2}-2\zeta^{-2})\,\mathrm{d}\zeta\right)^2 \qquad (1)$$

where $\zeta = 1+z$; and ζ_1 to ζ_2 is the range of redshift considered for the clusters. The covering fraction f is enhanced by the clustering of clusters[3] which, with scaling to $z\sim0.5$, corresponds to a factor of $19/\theta$, where θ is in arc minutes. Then if $\theta \approx 2$ arc min we have an f of 10^{-5}. This implies that there would be about four double images in the sky if there are more than 10 appropriate background quasars per square degree. We have taken $z = 0.5$, $\zeta_1 = 1.25$ and $\zeta_2 = 1.75$ so the values are appropriate to quasars of redshift $\sim$1. The velocity dispersion requirements are lessened if the clusters image the more numerous background quasars with redshifts of between 2 and 3. Including Abell richness class 1, we estimate an f of 5×10^{-5}. Consequently, there are likely to be $\sim$200 wide quasar images in the sky if the background quasar density is as high as 100 per square degree.

Bahcall and Soneira[3] remark that the clustering of clusters increases with richness so we may expect that it is most pronounced for the more massive objects required here. The above estimate involves a large extrapolation of their correlation function to small scales (it was measured on more nearby clusters with scales of degrees instead of arc minutes). There are, however, a reasonable fraction of nearby clusters observed to be double[4,5] with the separation required by a 2 arc min lens image. An excellent candidate is the A399/401 cluster system which has velocity dispersions[17] of 1,424 and 1,294 km s^{-1}, respectively (both with an uncertainty of $\sim$400 km s^{-1}). The apparent separation of these clusters at $z = 0.073$ is 33 arc min, which corresponds to 4.8 arc min at $z = 0.5$. Simple geometry then indicates that the probability of observing them with a separation of 2 arc min is $\sim$0.1. For our second method, we now assume that the A399/401 pair is not unique and estimate its spatial density to be approximately one per volume corresponding to $z = 0.073$. There are then $\sim$70 pairs aligned to within 2 arc min over the sky out to $z = 0.75$. Again, we expect about four double quasar images in the sky from 10 background quasars per square degree.

The timescale over which double clusters may last with such small separations of a few megaparsecs is unknown, but it can take several crossing times (each $\sim10^9$ yr) for the clusters to merge[18]. The Coma, CA0340-538 clusters and other binary clusters are large clusters containing two giant galaxies and could well have been A399/401-like pairs at redshifts $z\sim0.5$. Indeed, the hierarchical clustering of clusters strongly suggests that double clusters were more common in the past.

We have ignored the possible brightness enhancement of the background quasars by the gravitational lens, which depends on the precise light paths. Any Sunyaev-Zeldovich[19] effect due to hot gas in the clusters is a factor of two smaller in a double cluster than from a single very massive cluster (and possibly even less if the beamsize is not matched to the cluster size). No single cluster has yet been reported with $\sigma_{\parallel}$ of 2,000 km s^{-1}. Our estimates suggest that several quasars at $z \gtrsim 1$ are gravitationally lensed into double images of 2 arc min separation by double clusters. The discovery of these images will enable the clustering of clusters to be pursued to distances well beyond that of the Abell catalogue.

Received 16 June; accepted 10 July 1986.

1. Naranan, R., Blandford, R. & Nityananda, R. *Nature* **310**, 112-115 (1984).
2. Paczynski, B. & Gorski, K. *Astrophys. J.* **248**, L101-L104 (1984).
3. Bahcall, N. A. & Soniera, R. M. *Astrophys. J.* **270**, 20-38 (1983).
4. Ulmer, M. P. & Cruddace, R. G. *Astrophys. J. Lett.* **246**, L99-L103 (1981).
5. Forman, W. *et al. Astrophys. J.* **243**, L133-136 (1981).
6. Hazard, C., Arp, H. & Morton, D. C. *Nature* **282**, 271 (1979).
7. Turner, E. *et al. Nature* **321**, 142-144 (1986).
8. Gott, J. R. *Nature* **321**, 420-421 (1986).