Volume 10 Number 12 1982 Nucleic Acids Research

The complete sequence and coding content of snowshoe hare bunyavirus small (S) viral RNA species

David H.L.Bishop, Keith G.Gould*, Hiroomi Akashi and Corrie M.Clerx-van Haaster

Department of Microbiology, The Medical Center, University of Alabama in Birmingham, Birmingham, AL 35294, USA, and *Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received 4 May 1982; Accepted 19 May1982

ABSTRACT

The complete sequence of the small (S) viral RNA species of snowshoe hare (SSH) bunyavirus has been determined, principally from a DNA copy of the RNA cloned in the E.coli plasmid pBr322. The viral S RNA (negative sense strand) is 982 nucleotides long (3.3 x 10^5 daltons) with complementary 5' and 3' end sequences. It has a base composition of 30.5%U, 25.8%A, 24.9%C and 18.7%G. In the viral complementary (plus sense) strand there are two overlapping open reading frames initiated by methionine codons. One reading frame codes for a 26.8 x 10^3 dalton protein, the other for a 10.5 x 10^3 dalton protein. The larger gene product is presumably related to the viral nucleoprotein (N) that is coded by the S RNA (Gentsch and Bishop (1978) J.Virol. 28, 417-419). The smaller gene product is probably related to the recently identified S RNA coded nonstructural protein (NS_S) induced in virus infected cells (Fuller and Bishop (1982) J.Virol. 41, 643-648).

INTRODUCTION

Snowshoe hare virus is a member of the California encephalitis serogroup of mosquito transmitted viruses (Bunyavirus genus, family Bunyaviridae, reference 1). Genetic and molecular studies (2) have established that bunyaviruses have a tripartite, negative sense, RNA genome consisting of a large RNA species (L, approx. mol. wt. 3 x 10^6), a medium size RNA species (M, approx. mol. wt. 2 x 10^6) and a small RNA species (S, approx. mol. wt. 0.4 x 10^6). It has been demonstrated that the lipid enveloped bunyaviruses have two external glycoproteins (G1, G2, approx. mol. wt. 110 x 10^3 and 35 x 10^3 respectively) required for virus infectivity (3); these glycoproteins are coded by the viral M RNA species (4). The three internal, circular, nucleocapsids of bunyaviruses (and the similar Uukuniemi virus, Uukuvirus genus, see references 1 and 2) are each composed of an individual species of RNA (L, M or S) plus multiple copies of a 20-25 x 10^3 dalton nucleoprotein (N) in association with a few copies of a 180 x 10^3 dalton large protein (5) that is believed to be a transcriptase component (1). The N protein is coded by the S RNA (6,7) as is a 7 x 10^3 dalton nonstructural protein (NS_S) that, by

0305-1048/82/1012-3703$2.00/0

tryptic peptide analyses, is unrelated to N (8). It is presumed that the L RNA codes for the 180-200 x 10^3 dalton protein (2).

In order to begin to investigate the coding strategy of bunyaviruses we have sequenced the S RNA of SSH virus. The sequence has been obtained in part from direct RNA sequencing of 3' end labelled viral RNA (9,10) as well as by specific oligonucleotide primer directed cDNA transcription using chain terminating dideoxynucleotides (10) and from a DNA copy of the S RNA cloned in the E.coli plasmid pBr322. The data obtained indicate that for the S RNA of SSH virus there are two overlapping open reading frames initiated by methionine codons in the viral complementary (plus sense) strand. The evidence that the products of these open reading frames correspond to the N and NS_S proteins is discussed.

METHODS

Sequence analyses of SSH S RNA by cDNA transcription using oligodeoxynucleotide primers and chain terminating dideoxynucleotides

The procedures employed for dideoxynucleotide sequencing have been described (10-12). Two oligonucleotide primers were used to obtain sequences representing the 3' half of the S RNA. One (5' dAGTAGTGTAC) was used to obtain sequences corresponding to residues 11-320 (see Fig. 2 and reference 10). The other (5' dTCTTCTTCCT) was used to obtain sequences corresponding to residues 211-400 (see Fig. 2). The latter oligonucleotide was synthesized using the Vega automated oligonucleotide synthesizer, a protected T residue (representing the 3' nucleoside of the primer) linked to a polystyrene support (13, Vega Biochemicals, Tucson, Ariz. USA) and precursor protected T and C deoxymononucleoside-3'-O-chlorophenylphosphorotriazolides (14).

Cloning of a DNA copy of SSH S RNA into the E.coli plasmid pBr322

For the specific transcription of the S RNA of SSH virus an octadecanucleotide was made (5' pAGTAGTGTACTCCACTCG). The oligonucleotide was synthesized by a solid phase phosphotriester method using a protected G residue linked to a polystyrene resin (Vega Biochemicals) and precursor protected mono- and dinucleotides (15-17). After uncoupling and deprotection, the oligonucleotide was purified by HPLC chromatography, and phosphorylated at its 5' terminus by polynucleotide kinase. The primer was used to make a cDNA transcript of SSH S RNA (17). Prior to cDNA synthesis, 20 µl of water containing 20 µg of SSH viral RNA and 10 µg of the phosphorylated oligo-

nucleotide primer, were heat denatured at 100°C for 2 min, adjusted to 175mM KCl and 125mM Tris-HCl (pH 8.0) and annealed at 50°C for 15 min. The mixture was then diluted into a 0.1ml (final volume) reaction containing (final concentrations) 70mM KCl, 50mM Tris-HCl (pH 8.0), 10mM $MgCl_2$, 500 μM each of dCTP, dGTP and dTTP, 500 μM (α -^{32}P) dATP (specific activity 400 mCi per mMole), 10mM DTT and 10 units of reverse transcriptase and incubated at 42°C for 2 hr. After phenol extraction and passage through a 1ml Sephadex G50 column, the products were alcohol precipitated, denatured at 100°C in 80% formamide and resolved by 3% polyacrylamide gel electrophoresis in 7M urea. A single band of product cDNA (estimated size of 1000 nucleotides) was identified by autoradiography. The band was eluted in 2M ammonium acetate, recovered by alcohol precipitation and the estimated 0.5 pmole of product tailed at its 3' terminus with deoxyadenylic acid residues. For tailing, the cDNA was incubated for 60 min at 18°C with 25 units of terminal deoxynucleotidyl transferase (P-L Biochemicals, Inc, Milwaukee, Wisc.) in a 30 μl reaction mixture containing (final concentrations) 90mM cacodylate-23mM Tris buffer (pH 6.8), 2mM DTT, 1mM $CoCl_2$, 17 μM (α -^{32}P) dATP (specific activity 1.6 Ci per mMole) and 150 μg of bovine serum albumin. From the decreased electrophoretic mobility of an aliquot of the product resolved in a 3% polyacrylamide-7M urea gel, it was estimated that essentially all the cDNA had increased in size by some 30-50 residues. The rest of the tailed cDNA was phenol extracted, separated from reaction ingredients by chromatography on a 1ml column of Sephadex-G50 and recovered by alcohol precipitation. The product was backcopied in an 0.05ml reaction mixture containing 50mM Hepes-NaOH buffer (pH 7.0), 5mM $MgCl_2$, 25mM NaCl, 1.5mM DTT, 500 μM each of dATP, dCTP, dGTP and dTTP, 5 units of the "Klenow fragment" of DNA polymerase (Boehringer-Mannheim FRD), 0.1 μg of 5' phosphorylated oligo(dT)$_{12-18}$ primer (P-L Biochemicals Inc) and 0.5 pmole of product DNA. After incubation at 18°C for 3 hr it was estimated from the increased electrophoretic mobility of an aliquot of product run in a 4% polyacrylamide gel that greater than 95% of the cDNA had been converted into double-stranded DNA. The product was extracted with phenol, passed through a 1ml column of Sephadex G50, then digested for 5 min at 37°C with 5 units of S1 nuclease in 30mM sodium acetate (pH 4.6), 50mM NaCl, 1mM $ZnSO_4$ and 5% glycerol to remove single stranded DNA. After extraction with phenol and passage through Sephadex G50, the product was incubated for 15 min at 18°C with the "Klenow fragment" of DNA polymerase as described above (minus primer). The final double stranded DNA was then blunt-end ligated into the unique Pvu II site of the E.coli plasmid

pBr322 and used to transform competent E.coli MC1061 cells (18) as described previously (19). Of 200 colonies that grew on ampicillin-containing agar plates, 23 SSH positive clones were identified by Grunstein-Hogness hybridization (20). For hybridization short copy ^{32}P-cDNA was prepared by reverse transcription of total virion RNA (5 µg) with the 18-long primer (see above) under conditions where (α -^{32}P) dATP (20 µCi at 3000 Ci per mMole in a 20 µl reaction) was the sole source of dATP. One clone (S17) was used to obtain plasmid DNA for sequencing.

Sequencing of the SSH S17 clone

Sequencing was carried out by the Maxam-Gilbert method (21) using the formic acid protocol (22) for the A + G reaction. The strategy employed for sequencing the clone is depicted in Fig. 1. In toto approximately 93% of the sequence was obtained from analyses of both strands of the cloned DNA.

RESULTS

RNA and dideoxy sequence analyses of SSH S RNA

Partial sequence data for SSH S viral RNA species (residues 1-239) have been reported (10) from analyses involving chemical degradation of ^{32}P-pCp 3' end-labelled RNA, as well as from dideoxy sequencing using a decanucleotide primer (5' dAGTAGTGTAC) that is complementary to the 10 terminal 3' nucleotides of the viral RNA (9). A second decanucleotide primer (5' dTCTTCTTCCT, equivalent to residues 201-210, see Fig. 2) was synthesized and,

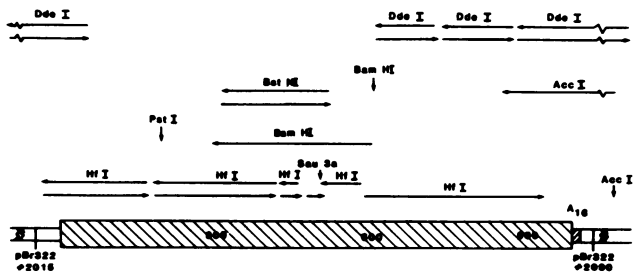

Fig. 1. The sequencing strategy employed to analyse SSH S clone 17. The restriction enzyme digestion products (enzyme names over the arrows e.g. HfI = Hinf I) used to sequence the clone were either strand separated (arrow pairs of opposite polarity), or recut with an alternate enzyme (vertical arrow in line with a horizontal arrow e.g. Acc I product recut with Bam HI). The orientation of the viral complementary DNA in the pBr322 vector is indicated as is the stretch of the 16 adenylic acid residues at the 3' end of the clone presumably originating from the cDNA polyA tailing (see Methods).

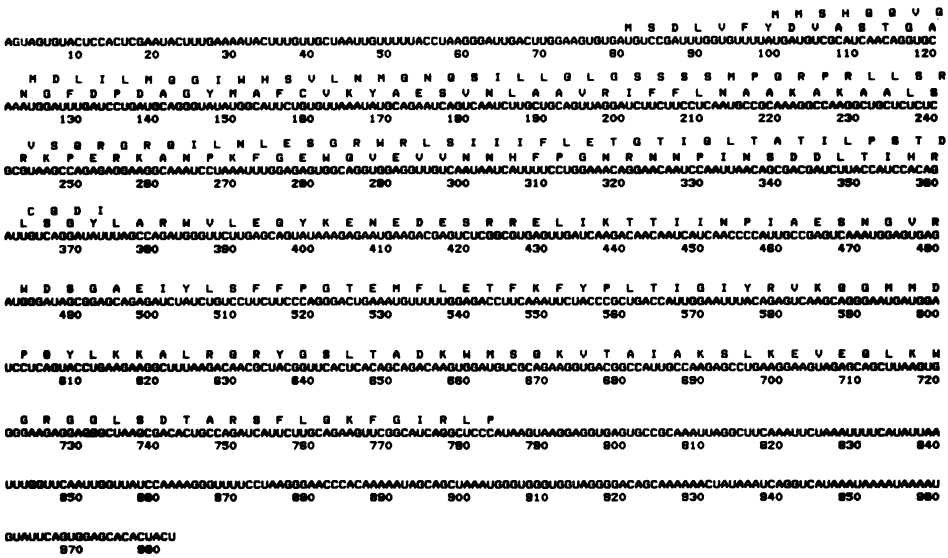

Fig. 2. The nucleotide and amino acid sequences deduced for SSH S viral RNA species. The RNA sequence, written as the complementary (plus sense) strand, was deduced from analyses of a DNA copy cloned in pBr322 (S clone 17, see text), in addition to direct RNA and dideoxy sequence data (9,10). The nucleotides are numbered below their respective residues. Amino acids of polypeptides (coded by contiguous open reading frames and initiated by methionine codons) are centered over the corresponding nucleotide triplets.

together with confirmatory dideoxy analyses with the terminal primer, used to obtain additional sequence data (up to residue 400, see Fig. 2).

Cloning a DNA copy of SSH S RNA

The terminal 3' nucleotides of SSH S, M and L RNA species are homologous in sequence for 11-13 residues (9,10). Therefore in order to clone the SSH S RNA species an octadecanucleotide primer was synthesized (5' dAGT AGTGTACTCCACTCG) that is specific and complementary to the 18 terminal 3' residues of the S RNA species (9,10). Using this primer, a copy of the SSH S RNA species was cloned in the E.coli plasmid pBr322 (S clone 17). End-labelled restriction enzyme digestion products of the clone were strand separated, or recut with alternate restriction enzymes (see Fig. 1), and the products separated by gel electrophoresis prior to sequencing (Fig. 2). Nucleotide sequences spanning restriction enzyme sites were determined from analyses of the products of other enzyme digests as indicated in Fig. 1. The Hinf I overlap at residues 173-177 was not determined in view of the alternate data obtained by direct RNA and dideoxy sequencing (9,10).

Characteristics of the SSH S clone 17 and the predicted S RNA gene products

The orientation of the DNA clone in the pBr322 vector is shown in Fig. 1. The 3' end of the clone terminates in a stretch of 16 adenylic acid residues presumably derived from polyA tailing of the cDNA. The 5' end of the clone lacks the 3 terminal nucleotides shown in Fig. 2; they were probably lost during the S1 nuclease treatment. Other than these 3 residues, the analyses of the DNA clone confirmed the RNA and dideoxy sequencing information obtained previously (9,10).

From the sequence data it is deduced that the SSH S RNA is 982 nucleotides in length (3.3 x 10^5 daltons) with a viral RNA base composition of 30.5%U, 25.8%A, 24.9%C and 18.7%G. The ends of the RNA are complementary in sequence and, depending on the arrangement, capable of forming hydrogen bonded structures as depicted, for example, in Fig. 3.

The non-coding region at the 3' end of the viral RNA is 79 nucleotides in length. At the 5' end the non-coding region is 195 residues long. The first open reading in the viral complementary sequence (negative sense) that is initiated by a methionine codon begins at residue 80 and codes for a primary gene product that is 235 amino acids in length (approx. 26.8 x 10^3 daltons). The amino acid composition of this gene product is given in Table 1. It has a calculated net positive charge at pH 7.0 of +6, assuming glutamic and aspartic acid are each -1, and arginine and lysine each +1, with histidine assumed to have a +0.5 charge at this pH.

A second open reading frame in the viral complementary sequence is initiated by 2 methionine codons (residues 99-104) and codes for a primary gene product that is 92 amino acids in length (Table 2, approx. 10.5 x 10^3 daltons). This second open reading frame overlaps that of the other reading frame and codes for a protein with a net positive charge at pH 7.0 of +3.

The longest stretch of open reading in the third reading frame of

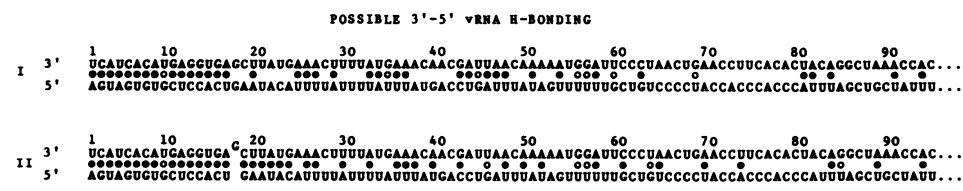

Fig. 3. The complementary end sequences of SSH S viral RNA. Two possible arrangements of hydrogen bonded 5' and 3' end sequences of the SSH S viral RNA are given.

Table 1. Amino acid composition of the 235 amino acid gene product

Ala	(A)	21	Leu	(L)	20
Arg	(R)	15	Lys	(K)	18
Asn	(N)	13	Met	(M)	6
Asp	(D)	12	Phe	(F)	14
Cys	(C)	1	Pro	(P)	10
Gln	(Q)	8	Ser	(S)	14
Glu	(E)	15	Thr	(T)	10
Gly	(G)	16	Trp	(W)	5
His	(H)	2	Tyr	(Y)	10
Ile	(I)	12	Val	(V)	13

Total number of residues = 235; MW 26,768

the viral complementary sequence extends from residue 529 to residue 714 and codes for 62 amino acids. There is also an open reading frame extending from residue 828 to the end of the RNA. However both these sequences lack methionine codons. In the viral (minus sense) strand, the longest open reading sequences initiated by methionine codons code for products of only 51 and 46 amino acids.

DISCUSSION

Genetic and molecular studies of SSH virus have established that the S RNA codes for the viral nucleoprotein, N, as well as a nonstructural

Table 2. Amino acid composition of the 92 amino acid gene product

Ala	(A)	1	Leu	(L)	14
Arg	(R)	7	Lys	(K)	0
Asn	(N)	3	Met	(M)	6
Asp	(D)	3	Phe	(F)	1
Cys	(C)	1	Pro	(P)	3
Gln	(Q)	12	Ser	(S)	12
Glu	(E)	2	Thr	(T)	5
Gly	(G)	5	Trp	(W)	2
His	(H)	2	Tyr	(Y)	0
Ile	(I)	10	Val	(V)	3

Total number of residues = 95; MW 10,478

protein NS_S (2,6-8). A review of the S RNA sequence reveals that in the viral complementary RNA there is only one contiguous open reading frame of sufficient size (235 amino acids, 26.8 x 10^3 daltons) to code for N. The N polypeptide recovered from virions, or from infected cell extracts, has an estimated size in SDS polyacrylamide gels of 21 x 10^3 daltons (23). There is no available information on the sequence of SSH S mRNA species so that it is not known whether there are spliced S mRNA species derived, as in the case of certain mRNA species of DNA viruses, or influenza orthomyxovirus, from non-contiguous regions of the genome. However, it has been shown that, unlike influenza virus, bunyaviruses do not require a cell nucleus for developing a productive infection (24), nor is the synthesis of their mRNA species particularly sensitive to actinomycin D, or α-amanitin (25). Preliminary amino acid and tryptic peptide analyses of SSH N protein agree with the postulate that N represents the gene product of the 235 amino acid open reading frame (F. Fuller and D.H.L. Bishop, unpublished data). However, whether the mature N protein has conserved the amino and carboxy terminal amino acid sequences of the primary gene product is not known.

The recent analyses which established that the SSH S RNA codes for NS_S, also demonstrated that N and NS_S do not share either arginine, or leucine, labelled tryptic peptides (8). Lysine, although an efficient precursor of N, does not appear to be incorporated into NS_S. By contrast, methionine is efficiently incorporated into both NS_S and N (F. Fuller and D.H.L. Bishop, unpublished data). The reported size of the SSH NS_S polypeptide is 7.4 x 10^3 daltons (8), significantly smaller than that predicted from the second contiguous open reading frame that starts at nucleotide residue 99 (Fig. 2). This reading frame predicts a product that is 10.5 x 10^3 daltons in size, with 6 methionine, 7 arginine but no lysine residues (Table 2). The NH_2 terminal 42 amino acid tryptic peptide of this product contains a long stretch of hydrophobic amino acids that are likely to render it insoluble. The primary gene product also predicts 6 other arginine tryptic peptides ranging from 2 to 9 residues in length. Only 4 such soluble tryptic peptides have been identified in digests of NS_S (8). Two of the 6 predicted peptides contain leucine residues, however analyses of NS_S have identified only 1 leucine labelled soluble tryptic peptide (8). Therefore if NS_S is the product of this gene, it is likely that it represents a processed derivative. Neither of the other stretches of contiguous open reading in the S viral complementary RNA (residues 529-714 and 829-982) predict polypeptides containing methionine residues, also both include

lysine residues and 2-3 tryptic peptides containing leucine. For these reasons it is unlikely that either sequence represents the gene coding for NS_S. Preliminary peptide sequence analyses of NS_S indicate that it has a tryptic peptide that has the sequence LLSR (F. Fuller and D.H.L. Bishop, unpublished data). The only place such a peptide is predicted from the primary RNA sequence is in the viral complementary RNA sequence in the second reading frame (i.e. residues 231-242, see Fig. 2). Therefore it appears probable that NS_S is a processed derivative of that reading frame although where processing occurs is not known.

If it is correct that N and NS_S are coded by overlapping reading frames then the question is raised of the identity of the mRNA species responsible for their synthesis. No direct analyses of the S mRNA sequences have been reported so that it is not known whether one mRNA species functions as a template for the synthesis of both gene products, or whether there are distinct mRNA species for each product. In this connection, it will also be important to determine exactly where translation of N and NS_S initiate. In addition to later positions there are methionine codons at residues 80-82 and 149-151 (first reading frame) and 99-101, 102-104, 123-125, 138-140 and 168-170 (second reading frame). Of interest to this issue is the possible

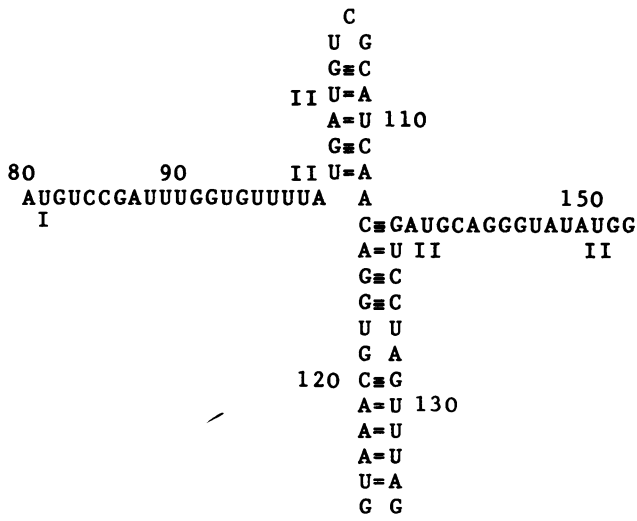

Fig. 4. Possible secondary structure of the region of viral complementary RNA containing the initial AUG codons in the first (I) and second (II) reading frames.

secondary structure of the viral complementary sequence between residues 99 and 137 (Fig. 4), the region which spans the initial AUG codons of the second reading frame. If both N and NS_S are synthesized from a single mRNA species, then whether or how NS_S synthesis is regulated will need to be determined.

The complementary nature of the end sequences of the viral RNA species of SSH S RNA, as noted previously for the related La Crosse bunyaviruses (26), is probably responsible for the circular nature of the viral S nucleocapsids (for a review of literature pertaining to the structure of the nucleocapsids of members of the Bunyaviridae family, see reference 2). It is likely that a similar situation exists for the M and L RNA species and nucleocapsids. To what extent the ends of the RNA are hydrogen bonded and what role the N protein plays in maintaining the circular nucleocapsid configuration, is not known.

ACKNOWLEDGEMENTS

We would like to thank Professor G.G. Brownlee for advice and help in the cloning of the S RNA. This work was supported by an MRC programme grant to GGB and by contract DAMD-17-78-C-8017 from the U.S. Army Medical Research and Development Command.

REFERENCES

1. Bishop, D.H.L., Calisher, C., Casals, J., Chumakov, M.P., Gaidomovich, S.Y., Hannoun, C., Lvov, D.K., Marshall, I.D., Oker-Blom, N., Pettersson, R.F., Porterfield, J.S., Russell, P.K., Shope, R.E. and Westerway, E.G. (1980) Intervirology, 14, 125-143.
2. Bishop, D.H.L. and Shope, R.E. (1979) in H. Fraenkel-Conrat and R.R. Wagner (Eds.), "Comprehensive Virology, Vol 14", Plenum Press, New York, pp. 1-156.
3. Obijeski, J.F., Bishop, D.H.L., Murphy, F.A. and Palmer, E.L. (1976) J.Virol. 19, 985-997.
4. Gentsch, J.R. and Bishop, D.H.L. (1979) J.Virol. 30, 767-770.
5. Obijeski, J.F., Bishop. D.H.L., Palmer, E.L. and Murphy, F.A. (1976) J.Virol. 20, 664-675.
6. Gentsch, J.R. and Bishop. D.H.L. (1978) J.Virol. 28, 417-419.
7. Cash, P., Vezza, A.C., Gentsch, J.R. and Bishop, D.H.L. (1979) J.Virol. 31, 685-694.
8. Fuller, F. and Bishop, D.H.L. (1982) J.Virol. 41, 643-648.
9. Clerx-van Haaster, C. and Bishop, D.H.L. (1980) J.Virol. 105, 564-574.
10. Clerx-van Haaster, C.M., Akashi, H., Auperin, D.D. and Bishop, D.H.L. (1982) J.Virol. 41, 119-128.
11. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci.USA 74, 5463-5467.
12. Zimmern, D. and Kaesberg, P. (1978) Proc.Natl.Acad.Sci.USA 75, 4257-4261.
13. Ito, H., Ike, Y., Ikuta, S. and Itakura, K. (1982) Nucl.Acids Res. 10, 1755-1769.

14. Miyoshi, K. and Itakura, K. (1980) Nucleic Acids Res. Symposium Series No. 7 281-291.
15. Gait, M.J., Singh, M., Sheppard, R.C., Edge, M.D., Greene, A.R., Heathcliffe, G.R., Atkinson, T.C., Newton, C.R. and Markham, A.F. (1980) Nucleic Acids Res. 8, 1081-1096.
16. Gait, M.J., Popov, S.G., Singh, M. and Titmas, R.C. (1980) Nucleic Acids Res. Symposium Series No. 7, 243-258.
17. Winter, G., Fields, S., Gait, M.J. and Brownlee, G.G. (1981) Nucleic Acids Res. 9, 237-245.
18. Huddleston, J.A. and Brownlee, G.G. (1982) Nucleic Acids Res. 10, 1029-1038.
19. Casadaban, M.J. and Cohen, S.N. (1980) J.Mol.Biol. 138, 179-207.
20. Grunstein, M. and Hogness, D.S. (1975) Proc.Natl.Acad.Sci.USA 72, 3961-3965.
21. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
22. Bishop, D.H.L., Huddleston, J.A. and Brownlee, G.G. (1982) Nucleic Acids Res. 10, 1335-1343.
23. Gentsch, J., Bishop, D.H.L. and Obijeski, J.F. (1977) J.Gen.Virol. 34, 257-268.
24. Pennington, T.H., Pringle, C.R. and McCrae, M.A. (1977) J.Virol. 24, 397-400.
25. Vezza, A.C., Repik, P.M., Cash, P. and Bishop, D.H.L. (1979) J.Virol. 31, 426-436.
26. Obijeski, J.F., McCauley, J. and Skehel, J.J. (1980) Nucleic Acids Res. 8, 2431-2438.