The complete nucleotide sequence of the RNA coding for the primary translation product of foot and mouth disease virus

A.R.Carroll*, D.J.Rowlands* and B.E.Clarke*

Animal Virus Research Institute, Ash Road, Pirbright, Woking, Surrey, UK

Received 10 January 1984; Revised and Accepted 16 February 1984

ABSTRACT

The complete nucelotide sequence of the coding region of foot and mouth disease virus RNA (strain A_{10} 61) is presented. The sequence extends from the primary initiation site, approximately 1200 nucleotide from the 5' end of the genome, in an open translational reading frame of 6,999 nucleotides to a termination codon 93 nucleotides from the 3' terminal poly (A). Available amino acid sequence data correlates with that predicted from the nucleotide sequence. The amino acid sequence around cleavage sites in the polyprotein shows no consistency, although a number of the virus-coded protease cleavage sites are between glutamate and glycine residues.

INTRODUCTION

Foot and mouth disease (FMD) is a highly contagious viral disease of clovenhooved animals and is of paramount economic importance. The virus (genus: aphthovirus; family: Picornaviridae) has a single stranded positive sense RNA genome of approximately 8,500 nucleotides. The genome is 3' polyadenylated, can serve directly as a messenger and contains a poly (C) tract (100-200 nucleotides long) located approximately 400 nucleotides from the 5' end (1).

Unlike most eukaryotic mRNAs, FMDV RNA is not capped at its 5' end but has a small viral-coded protein (VPg) covalently attached to the terminal 5' uridine residue (2, 3). The VPg protein is not required for translation and, by correlation with poliovirus is probably involved in replication (4, 5). The RNA of FMDV, like other picornaviruses, appears to have a long 5' untranslated leader sequence. Removal of the poly C tract of FMDV and all nucleotides to its 5' side has no effect on the spectrum of proteins generated by <u>in vitro</u> translation. This indicates that the site for initiation of translation is located to the 3' side of the poly (C) tract (6).

The initial translational product of FMDV RNA would be a polyprotein of about 250,000 daltons. However, this protein is not observed as it is nascently processed into the viral structural and non-structural proteins as shown in Figure 1.

We describe here the complete nucleotide sequence of the region coding for this poly protein from one strain of FMDV.

MATERIALS AND METHODS

Recombinant Plasmids

The construction of recombinant plasmids has been described previously (7). Briefly double stranded cDNA was synthesised using RNA from FMDV strain A_{10} 61 and inserted by G:C tailing into the Pst I site of pAT 153. Two of the resulting recombinants (pFA76 and pFA206) contained inserts corresponding to ≈85% of the FMDV genome. Detailed restriction maps of these recombinants were generated by standard procedures. pFA206 was subcloned into two more manageable size clones pFA206 α or pFA206 β . Initially pFA206 was digested with EcoRI, recircularised using T4 DNA ligase and used to transform <u>E. coli</u> MC1061 by standard procedures (8, 9). This resulted in the clone pFA206 α representing those sequences to the 3' side of the EcoRI site in pFA206 (see Figure 1). pFA206 β was generated by digestion of pFA206 with EcoRI and PstI, gel purification of the required band and ligation into EcoRI/PstI digested pAT 153.

DNA sequencing

All DNA sequencing was carried out by the method of Maxam and Gilbert (10). Recombinant DNA was sequenced using uniquely 5' or 3' 32 P labelled restriction fragments as described previously (11). Primer extension sequencing was carried out as described by Rowlands <u>et al</u> (12). Briefly, a 5' 32 P-labelled primer (Cell Tech. Ltd., UK) was used to direct the synthesis of cDNA using reverse transcriptase and viral RNA as a template. The resulting 5' labelled cDNA was size fractionated and sequenced by the method of Maxam and Gilbert (10).

Analysis of Sequence Data

DNA sequence was analysed using an Apple II microcomputer as described previously (13).

RESULTS AND DISCUSSION

Sequencing Strategy

The construction of cDNA clones from FMDV RNA (serotype A_{10} 61) has been described elsewhere (7). These clones represented in total about 85% of the viral genome (Figure 1). The nucleotide sequence of the region coding for the structural protein (VPs 1-4) has been previously reported (11). As shown in Figure 1 a large clone (pFA206) having a cDNA insert of 5.4Kb represented the major part of the genome coding for the non-structural proteins. To simplify sequence analysis this insert was sub-cloned into two halves (pFA206 β and pFA206 α) as described in methods. Detailed restriction analysis of these clones (see Figure 1) allowed a sequencing strategy to be carried out so that most regions of the genome were sequenced in both strands. Those regions which were not sequenced in both strands are clearly indicated in Figure 1 and these regions were determined from



<u>FIGURE 1</u> Organisation of the FMDV genome. The positions of the gene products is based on previously published work (1). The naming of the polypeptides has been done using both existing names and the recommendations agreed at the 3rd European Study Group on moleculear biology of picornaviruses, Urbino, Italy, September, 1983. The new nomenclature is indicated above the corresponding polypeptides and the old nomenclature is indicated below. The alignment of two cDNA clones with respect to the viral genome is shown. The Eco R1 site was used to sub-clone pFA206 into α and β as indicated. The sequence of 2820 nucleotides from the 5' end of the clone pFA76 has been reported previously (7). The sequence strategy is illustrated by the bars and arrows indicating the restriction sites used and the extent of the sequence obtained. Primer extension sequencing was used to obtain sequence to the 5' side of pFA76 and the location of the primer and sequence obtained is shown. The abbreviations used for restriction enzymes are: B, Bam HI; D, Hind III: E, Dde I; F, Hin FI; G, Bgl II; H, Hpa II; L, Bgl I; M, Sma I; P, Pst I; R, Eco RI; S, Sal I; T, Taq I; V, Pvu II; X, Xho I.

several independent sequencing experiments.

Previous mapping studies using defined T_1 oligonucleotide probes had indicated that the clone pFA206 did not include sequences extending to the poly (A) at the 3' end of the genome (7). The sequence data show however that this clone does in fact contain the whole of the 3' end of the genome including 33 bases of the poly (A) tract.

None of the cDNA clones hybridised with T_1 oligonucleotides derived from the 5' end of the genome suggesting that this region was not represented in our clone banks. In order to obtain 5' end sequence beyond the region represented in the clones a primer extension method was employed. A synthetic oligonucleotide (Cell Tech Ltd., UK) whose sequence was derived from the 5' proximal region of pFA76 (see Figure 1) was used to prime cDNA transcripts which were sequenced as described in methods. Approximately 450 bases were sequenced using this primer.

2021	E GS	A Room	84 SSr	6 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	55 55 55 55 55 55 55 55 55 55 55 55 55	883 293 293	1yr Tyr	8 <u>5</u> 5	ASD ASD ASD	MET NUC	1440 GCC Ala	1560 Thr	A200	85.2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
Ly s	웧놑	Br	SF	Lys Lys	Eg	500	A B B	Lys Lys	ົຮີອ	AAC A	, SSA	CUC		Age
ANG Lys	Ee	Ala Ma	83	Lys Lys	85	AC	CUC	Asp	CUC	PP PG	SF	Egg	Phe	AUU
GAA	AAU Asn	AC AC	Asp	35	ACA	Page B	500 V	장권	Ala Ma	AGA	ASD	58 S	Acc	550
C17	CAC CLU	500	Ser	GUU Val	Ser	Ala	CUC	AcA	<u>B</u> r	ប្លដ្ឋ	Ser	5SS L	Sc	Ser
105 AAC Asn	222 2 2	345 AUC Ile	Acc Thr	555	ŘET Č	ក្លីភ្លឺន	See S	Sol t	CUA Val	ည်း ကို	CUC CUC Val	1545 UAC	QS SS SS SS SS SS SS SS SS SS SS SS SS S	Leu Age L
CAU His	8s S	00CU 017	GUC	ACA	Ser D	Ser	В°	CAC Asp		L1e AUC	_85	์ ซิร์	ALa	Tyr
Leu	AAC	Acc	Š	AAA	AAC	Ala	55	Phe	CUU	Phe	CCC	GAC	GAA	ACA
ACA	Tyr	Glad	Ala	Trp	S.C.C.	Leu	Acc	g	GAU	65G	ខ្លួន	GCA	ALa ALa	ABC
Phe Phe Phe	GUC	Lee	Phe	GAC	Tyr	AAA Lys	Acc	Phe	SUL	CAC	Ser D	ςς ΩC	CUC	Ser
CLE 0	210 170 170 170	ဦးရေ	200 Va1 Va1	520 520 520	CAA CAA	Ser A	See S	Lys AG	600 617 617 617	Pro 200	Leu Alo	530 CCA	650 CAC Asp	NET OF
Mere I	CAC Asp	CAC	GCA	AAU	Gig	Phe	Thr	Lys	Asc	Pac_	CUA	CUU	[DU]	Lac CAC
AAG	Phe	AAA	CAC	Leu	MET	Trp	Acc	Phe	ACC	Eeu	MET	Pro	Leu	AAA
CCA CLY	Phe	AUC	GAU	Pro	Tyr	GAC Asp	CAC	Phe	MET	Acc	Cuc Val	Pec 3	AAC	GCA Ala
AAA Lys	BS	AAC	S.G	GAA	Tyr	AAC	550	ACA Arg	Tyr	Leu	CUU	S ≣a	ACA	GCC
Aco Thr	CAA CAA	Trp CCC Trp	CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA	S55 CMA G1n	ASD ASD	ASC ASC	915 Asn	636 GNC GNC	155 CCA Ala	610 610 610	1883 1883	583	Phe	Z
AGA Arg	CAU Asp	AUC	AAA Lys	GAU	AAC	GIN	Arg Arg	Lag Ala	_UAU TYT	L'AN	Thr	550	្កែខ្លួង	L Ser Ser
Ser	GUC	CUC	MET	Tyr	AUU Ile	Acc	Thr	GIn	Ser	AAA	Trp	AAA	550	GUC Val
Arg	Tyr	Leu	Phe	ខ្លះ	AUA Ile	AAC	Thr Thr	GUC	Asp	GAA	D S S S	Ser	5 L	GAU Asp
Phe	ACC	Ala Ala	AUU	GUC Val	Ser	Acc	Lau	CUC	CUC	Arg	AAA	Pro CC	UAC	Phe
du Leu	Phe Phe C	SS L	600 61 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	Pres	660 617 617	ACA ACA	AUC AUC	Arg 800	Egz	ACA ACA	380 CAC	Leu Cgo	AAC ASD	Lys AAC
Acc	UUC Leu	Pro	GCA	GUA	ACU	His	Ar 8 Ar 8	_ Å F	CAC	CAC_	Lys	້ອີວິ	Thr	GCC
AAG	Cin	CCA CCA	CAC	Leu	AAC	ACA	GAC	GAG	583	Phe	AAG Lys	CLY CLY	AAG	Leu
AUU Ile	Leu	5 C C C C C	Phe	GUC	550	Ser	GAA	UUA Leu	Phe	ALa ALa	Tyr	Ala	20 20 20 20 20 20 20 20 20 20 20 20 20 2	Cau
GAG	AUC	C C C	GAU	GAC	Ser	P AC	Leu	GCC	GUC	AAG	GGG	CUC	85	Arg
AGA AGA	165 Thr CCC	285 CAC His	Abge	se Sco	CHC CHC CHC CHC	ACA Thr	ဦး ကြိ	Soo s	5001	17 Trp CC CC Trp CC Trp	Asp CAU Asp	His His	AAC Asn Asn	Thr
AUC	AAC	Leu	Leu	Pro CC	AAC	GAC	Act	ACA	Lis	[TYL	CUU	TYN,	GAC
Leu	Leu	C.M.	Š	GAC	CAC	A B R	Thr	AAC	CAC	Proc	Arg	UAC	CUC	GAC
UAC	50 D d r	Leu	MET	ខ្លះ	В.S.	Ser	GAG	5 Loc	GAU Asp	CUC	AAC	Acc	AAG	GCA
GUA Val	S S C C	550	GAC	Thr Thr	550	SSC	GAA	550	Acc	MET	CUC	Proc	CIV CIV	Arg
red Edd	150 AAC Asn	Acc	88 F	120 120 120 120 120	ACC ACC Thr	750 61 u	7900 1900 1900	Argo 600	505 202	230 660 Ala	550	470 GCC Ala	TYT COST	
Ala	GAC Asp	Phe	550	85	GCA	Ash	LANC N	Cuu	ຼີອີງ		์ ฏิจู	Lie AUU	Val	Val
AUC	CAC	AAC	Asp	Tyr	Pro	Ser	A Se S	His	010 010	ga	Tyr	AAC	Br	Cuc
Phe	AAC	GAA	GUA	Pre	A Se s	200	ASA	GAU Asp	gg	Egg	ge	ALa	Asp	Tyr
SS UCΩ	AAU	Leu	SP MAC	Asp	S S S	50C	Ala C	Side A	AAA Lys	ςς ΩC	CuA	Tyr	P.S.	ge
ASN ASN	102 102 102 102 102 102 102 102 102 102	555	500 202	550 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	550 1900 1900	Ser	င်ငိုင်	010 0233	PP285	5002	200 L	COC Val	Thr Thr	č¥%
Thr	Arg	AGC VC	CUC	GAC Asp.	- 83	AUC Ile	Be	BF	ີ່ອີງ	ີວິດ	Lie N	- AG	Lys SA	83
Pro So	Ser	AUC	GAG	GAC	P.S.	UPC P	Ala	S S S S	Tyr	ASD	His CA	Dec.	85	Asp
¶ ¶ ¶	Tyr	A ^c Q	Ser	NUC 11e	500	Ash Ash	560	JAK D	55 SC	BE	Ala Ala	8g	Asp	Asp
<u></u> 월달	7	CAU Asp	ပ္လင္ရွိ	Ala Ma	ANG	Asp	Pec	83	Be	33	B P	S.	ACA	Puc

1920 ACA	Ala Ala Ala		11e 11e	Asp	င် ရွှေ	PACK PCF	88ª₽	Va1 Va1	8500 1000 1000	SS a	zer	မ္က်င္လိုင္ရ	LYAG ARG	A200
CAC_	. 8 ⁵	ឹទ្ធខ្ម	Be Be	Ser	ACR.	. Sa		Asp. C	58°	al C	"Bg	TIe NO.	"Non	val
ខ្លះ	GAU Asp	Leu	550	Phe	84 14	CUC	Lys Lys	Mag	렇는	Ba	CIV	Ala	E.g.	Val
មិត	Ser	Arg Arg	GUC	UAC	Phe	Arg Arg	AUA	Bec Bec	N ² C	Cuc	Pac Bac	De C	AUC	Lys NA
A5 A5	ALa ALa	Leu	GAC	Tyr	85 5	A20 A20	Ala	Phe	Bu	S S	Ser	AUC	580	ΩČ C
000 010 010	ACC Thr		ACC ACC Thr	Thr	2505 CCA A1a	ALa BCC BCC BCC BCC BCC BCC BCC BCC BCC BC	ES3	Prov	C.MS	Asp Asp Asp	KS &	Asp Asp Asp	နိုင္ပ်ိဳင္ရွိ	င်င်စို့
CUC	TYL UNIT	Phe	CAC	Ala	ANG	AUC	ີ ເອີ	~ပ္ပင္ရ	C.G.C.	ر ۲۹۳	်င္လီရီ	A		
560	Ala Ma	GAC Asp	CAC	CCA Ala	AAC	Ser	Proc	83	Be	Ser	ដ្ឋ	AUU 11e	Eug	600 Ala
BS	Tyr	AAA Lys	Arg	Arg	Tyr	8 <u>6</u>	A S S	Leu	A ² G	88 88	Ala	Asp	Asp	AUU
82 82	C AU Asp	000 01 01	ACA	Leu	600 Ala	SJ	S 5	AAC	Ser	Ala	8£	Arg	8 문	CAC
1890 100 11e	A500	ALa ALa	C20	2370 DUA Leu	ACU Thr	CGAC Asp	200	နိုင်ငံ နိုင်ငံ	Val	Ares Ares Ares Ares Ares Ares Ares Ares	210 CUC	ALACCA	цеб5	570 CUC Val
Tyr	, Scc	Ser C.	CUC,	B SS A	ិប្លខ្ម	ີຮູລິ	Tyr	C.G.C.	Leu Leu		HIS CAC		_ PAC_	ABC ABC
Ala	Ser	Age	GIn	C.C.	AAC Asn	Ser	Eg	Val	88	Ala	Phe	Leu	GUC	550
CUC	CUC	Ser	AcA	CUC	Ser	ACA	C NC	Asp	AAC	Ala	Eg	55	2 B G G G G G G G G G G G G G G G G G G	Ser
MET ACC	Tyr	GUC	GAU Asp	AUC	ACC	Ser	Ala	CI,	Phe	SE	Ser	AAA	Lys Lys	AAG
1875 UAC Tyr	5 <u>5</u> 22	Leu Leu	533 5335 5335	555 555 555	AAC Asn	S295 Asp	285 785	A COS	Asp Asp	άζ δ	Sci 25	Gid Gid Gid Gid Gid Gid Gid Gid Gid Gid	656	Lee Cor
Arg.	, AUC	[[GC]	ීපිටි	His CAC	Ser	Ser	AS	د دور	~ SS 5	្លតិខ័	ိည္စဥ္	Ala Ala	CIU GIU GIU	ົວເຄື
GCC	Ser	Acc	UAC	AAA Lys	ClC	6CA Ala	Mec	AAG Lys	5S S	Ba	Ser	ACA	Ser	Ala
AAA	Phe	GAC	AAC	CAC	GCC	Ser	858 860	Egg _	CAC His	Arg Arg	GAC	Cie Cie Cie	Ala	55
Ser	Acc	AAU Asn	GAG	Acc	GCA	Tyr	CUC	D 52 CUA Leu	Lys Lys	Ser	Ser	Leu	AUU Tie	Arg
1860 CAC Asp	200 200 200 200 200 200 200 200 200 200	GIU	COC CUC Val	CAA CAA	C1u C1u	2580 Lys	ဒိုင်ခြို	Asp CAC Asp	14CA	Leu 200	80.08 80.08	Asp CAC	TTC DC DC DC DC DC DC DC DC DC DC DC DC DC	A18055
Thr	AAA	F age	Thr.	MEC	ີ່ບູດ ມີບ	Asn	ິຊີ	"Page	Ser	Lee C		Ctu Gtu	ALa CC.	ABC
Ser	Ser	AAG	Acc	Leu	66C Ala	ACA	6AC CLu	ASC ASC	M	AAG	Lys	Sc	AAG Lys	GAC
00C 01V	AAC	560	Acc	GAC	GCU	550 17	CAC	Luc _	GAC	AUC	CUC	AcA	AUU	Leu
ACA	Eeg	CAC	GUC	AUU Ile	AAC Asn	GAC	AUC	Leu	GAG	Leu	GUC Val	Ser	SF	SF
1845 00C Phe	000 017 017	ACA Thr Thr	ស្តីខ្លួះ	2325 CUC Val	Pro	Tyr	ALa 600 800 800 800 800 800 800 800 800 800	800 2010 2010	555	LYS BE	Phe CCC	888 x	Asp Asp Asp CAC	552
MET	ACA	11e AUU	GAC.	CAU,	CUC		ີ ອີ ອີ	Lys N,		_ TYL	<u>T</u> ag	പ്പെട്	ိမ္မန္မ	۲, KG
DUC	GAC	GIn	6CA Ala	ACA	D d r D	Thr	ALa ALa	Mag	35	5 D L	Ser	Pec	AUC Lie	Ala
CAC	Trp	UAC	Ser	2 L L C C C C C C C C C C C C C C C C C	Acc	GCA	55	ដ្ដ	N SA ASIA	ដ្ដ	Asp	S S	Ala	A S S S S S S S S S S S S S S S S S S S
CCC	GAG	GUU Val	C.C.C.	Ser	Ee	E E E	AUC 11e	A a	AUC	Lys AM	Eee	Ser	Eeu	AAG
1830 Asn Asn	1950 Ala	ŝŝŝ	583 293 293 293	Egg	Asn all	2000 Va1	A18 A18 A18		50 20 20 20 20 20 20 20 20 20 20 20 20 20	ALa CCO ALa	150 150	ALa SCO ALa	E Seg	UAC Tyr
, DUC	CAC.	Cuc.	``₽₽	ິບະຈິ	ີ ອີອີ	"22" 420"	ີ ອີວິ		asp Asp Asp	ຼືອອີ	ິອີ		်းခြ	
<u>S</u> E	AUU	SF.	- SE	ASn	GAU	CAC H1s	Tyr	AAG	GUA	Asp	Leu	Lys AG	AAA Lys	550
ଞ୍ଚିଟ୍ର	S S C C	50 V		AUA Ile	CAC	Pro	Asn	CAA	Seg	Eec	55	Val	GUC	នួ
Ser	CAC His	55 C	달론	AAG	Arg	Aia Aia	Ba	Lys Lys	Lys Lys	55	Acc	Eug	E.G	Asp
1815 UAC	Ala SQ	Val Val	CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT	Val Val	Val Val	1900 Lange	င် ကိုင်သူ ကိုင်	Tyr Tyr	စိုင်ခြီ မ	Prop 1	Asp Asp	883 883 8	žä ^r	ASh Ash
35	Ala	Asn	Ppr VCV	Phe	cuu	TYr TYr	Υ ^α δα	Υς Υς Α	Phe	AGA Arg	်င်းရှိ	Alc	Geo Geo Geo Geo Geo Geo Geo Geo Geo Geo	ຼືອີງ
ACA	GAU	ACC	888 880	Arg 864	AUU 11e	ខ្លះ	<u></u> gr	Asp	AAC	AUC	53	BJ	SS	Asp
Tyr	GAU	ACA	ပ္လင္ရွိ	Asp	GAG	Eg	<u>B</u> g	S 2	Ser	Aga	읧	83 83	AAC	84
Tyr	ទ្ធះ	CAG	GAC	₩.	Leu	Ala	N	Ser	ACC	AM S	AUC Ile	Cuu	Lys AG	010 010

000 000 010	14 14 14 14 14 14 14 14 14 14 14 14 14 1	ACC ACC ACC	450 Asp Asp	4200 CUC Val	E CNC CNC CNC	탈명릴	1280 6760 6760	P200	800 800 800 800 800 800 800 800 800 800	A CUC	800 800 800 800 800 800 800 800 800 800	5160 610 610 610	5 28 28 28	
^오 분	Lys V3	Be	S G	WW WW	24	82	ă,	DG S	33	<u>د</u> لا	- Žď	ĒĒ	₩.	- <u>ပ</u> ိုင်
Be	BE	83	<u>E</u> e	39	53	č,	ŠĚ.	ŠŠ	25	<u>S</u>	934 84	N N N N N N N N N N N N N N N N N N N	95	ST.
H13	Asp	85 Br	A ^{CC}	ĔĔ	3£	3£	55	33	ĒĞ	0 V	D'O'	25	223	PCC Thr
AC	Lys Lys	TYL	Lys.	SE	3Å	252	A,	83	č¥		Ę.	ŠĚ	₹£	Asp
3735 Ser Ser	ଞ୍ଚିତ୍ରସ୍ତ୍ର	Eeco Eeco Eco	11 Perce	2022 2022	ŠS3	100	4545 CUC	Pro 4	AL CO	şgr.	ζ¥ζ	Sealer 1	P	eros Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Seco
11e	Asp	ASD	N N	85	COC	₹Š	8ª	83	<u>S</u>	Ly.	83	55	85	A SA ASA
GCA	ទុទ្ធ	Thr	GAC	Br	Ser	CUC	Asp	See.	CUC	AND Lya	CAC AS	ASC ASC	BE	MET
S 2	ASD	ACC	DD	Lys Lys	ANG	Page	Asp	AGA	Lys	Cuc	ĔĞ	PAC	H13	S S S S
6CA Ala	550	Ala	Lys MA	Phe	A 2	S S C S C S	CUC	610	5g	85	Ash Ash		ξ\$	GUC
န္တိရှိချိ	61 000 01 000 01 000	11e Miles	ASC ASC ASC		နိုင္လီဦးနီ			ACA ACA	₽ ₽ ₽	\$ 5 53	E GSS	A a c c c c c c c c c c c c c c c c c c	H Sol	5370 CUA Val
20C	E	AUC	AS ^D		<u>_</u>	_A22	Lys Lys	Be	۲. ۲	C.S.C.	C SC	ACA		"Bei
AAC	GAU	GUC	AUU	₩ ⁵	11e	32	85	55A	95	So Si	GUU	550	CAC Asp	CAC Asp
Ala Ma	GAC	Lys Lys	Lys Lys	₹ ³	В В Г	AUC Lle	Arg Arg	GUU Val	Arg Arg	₹ 8£	ខ្លះ	Asp	ACA	Lys VC
Ed	똜	Ser	Tyr	2 E C C C	AUC	ĘĞ	Lys Lys	AC	olo Glu	⊥§ਤੇ	ANG Lys	Eeu Ceu	CUC	UAC
005 Phe	205 Val	ASC ASC ASC	555	ACA ACA	500	2 B B B B C C	₹602 875	Ser Ser	Egg B	61u GM3	₹ S S S S S S S S S S S S S S S S S S S	£5€	Sec.	Acc
ົອະ	Cuu Val	"Dae Be	Asp CAC	ζÅς ΓΥς	ζ ΥG Γ	្លៃខ្ល	"AC	725 P	ិភ្លះ	ζÅς Έγει	AS ^A C		B AG	္ကိုမ္ခ်ီ
AAG	GUC	ខ្លះ	AAG Lys	M	Phe	Eg	Glu	CCC CCC	550	CUU Val	583	Lys Lys	555	ALa ALa
GUY	Acc	Lys Lys	Ala	GAA	AUU Ile	GAG	828 828	Ser	Ser 7	CUC	MET	GAC Asp	Arg 800	CAC
55	85	56	Ser	CUU	PcC	0.00 CGAC	AUC	Thr	PAC Para	ទួក	GUC	UAU TyT	CAC H1s	CIV CIV
5000	855 Break	SAS SAS		Ala	H1s CAC	Sec. 2	MECS	GIA GIA	FUL	A2000		No.	E G	Ser
ິສະ	AAC Asn	Cac Asp	Asp 4	_ ₩	"Ser Ser			้ ออา	ร ีรูรัฐ	Lys L	ΓÂΓ [™]	19 19 19 19 19 19 19 19 19 19 19 19 19	ົກສ	Jec.
Lys Lys	Tyr	GAG	AUC	SS	Ser	Ser	Cuc	ខ្លួះ	- Au	CUA	85	Ala	Suc	AUU
56	56	Leu	GAC	ASC	Val	GAC	AUA	ASC	Aa	AAA Lys	Leuc	Phe	CAU H1s	Leu
Arg	GAC Asp	Ser	Phe	Lec C	AAA	CAC His	ASD	Lys Lys	S PACE	GUA	Asp	Ee	₽ <mark>₽</mark>	AGA
E SS	Phe C	800 P	His CAC	년 등 관	CTC CTC CTC CTC	365 CUA Val	\$G4	600 1000 1000	SS L	AAA Lys	Thr 2005 rd	H13 H13 H13	555	500
ິບຈິ	CAC His	₩ E E C C	Pac *	T ^a CC [*]	HIS CAL	₽¥ [®]	້ ອິ ລ	⁷ 88	1ี่¥ี่5ื	້ອງ	ឹម្លខ្ល	្លឹងខ្ល	ິວສ	Cuu
CUU	Asp	ដ្ឋ	Arg	3 SC	Eg	550	Leu G	GIU	Asp	ម្លះ	S S S S S S	ទ្លដ្ឋ	ACG	GAU
Val	52	ម្លះ	Arg Arg	Asp	GAG	GAG	ACC Tar	Asp	Gluco	AAA Lys	Alac .	Val	Lec C	CCC Ala
CUC	Asp	AUC	ASD	TYC.	CUC	Phe	Eeu	E di	Ala	955		Leuc	Ala	AAC
88°5	۶ <u>5</u>	Pacc 83	883 288	295		Pecco Pecco			202	ALSO ACA	SCE 5	PAC ACC ACC ACC ACC ACC ACC ACC ACC ACC	83 H	ASI
Cta Gta	്ഋൂ		_ ຊີຊີ	Pace *	GA &		دوم ا	Lys,	້ ອີ້ອີ	10°	55	Å 80	56G	Cuc
Sc	Še	8 E	Asp	M	I)e	Ile NUC	Ala	Asp	ALa	E SE	Br	2 E E	AC	CUU Val
ACA	TYL	로	Se	Ala	SE	N.S.C.		SAC	SE	82	al Cuc	SA	22	C1y C1y
Ser Ser	35	85	Š	35	5 Sec	5gg	Ne la	월놀	910	56		2 E E	A.	CUU
Lys PAG5	Val Val	A COS	2023	SUL	502	Mon a	500 n	No I	Ser a	283 a	ਲਿਤੇ	555	202 Mal	505 505 Val
Ser	ິຍະອ	. ARC	្លែខ្ល	ANC.	Val.	HIS COL	- Be		10 N	N ST ST	ASD 4	3GA 5	N.S.	ំអូ
SE	ASP ASP	85	85	SA	E.G	8g	AAC	AN A	Sig		85	<u>S</u> E	1 and	А С
29g	ALC.	Ala	ACA	H15	85	Sg	Seg	5GA	Asp Asp	8	Ba	55	200	550
S S S S S S S S S S S S S S S S S S S	ACA	Ba	85	SAC A	TYL TYL	AMA Lys	AG Lys	A A A	Asp	000	AN S	S C	2 Be	Lys A

5490 GC AGU CAC UCU GCA GGU GGC ANU GCA GUU GCA UAC 1Y Thr His Ser Ala GIY GIY Asn GIY Val GIY Tyr	C5610 CC AGA GAU GUG GAA GAC GGC GUC CAC GUC AUC GGC hr Arg Asp Val Giu Giu Arg Val His Val HET Arg	5730 District and con could cut car car cut any UUC ev Asn civ ciy val val Lev Asp Asp Val Ile Phe	5850 AC ACC CUC CUC GGU ACC GCA ANU GCC CCA UUC ACC Is Ser Val Leu GJY Thr Ala Asn Ala Pro Leu Ser	AA CCC CCC CCU CCC CUC 5985 AA CCC CCC CCU CCC CUC CAC AAC CCC ys Arg Arg Cly Ala Leu Ile Asp Phe Clu Asn Cly	6090 MA AUU CCA CCC ANG CAA CUA CCC CCC CCC ANG LIU ILE ANG PRO MET CLU LyS VAL ANG AUG CLY LYS	6210 LAC AAC GGA CCA CAA ANU GGC UCU GGG GUC GGU UGC LAN ASN GIY Pro GIN ILE GIY SER ALA VAI GIY CYS	6330 Ca Nac cac ucc acu cac cod auc anc auc auc duu La Asn His Cys Ser Asp Ala Met Asn lle Met Phe	6450 Arc arc coc auc acu cub can coc coc auc cca thu asn lys arg ile Thr Val clu cly cly MET Pro	6570 100 GAG CAC ACU UAC ACC AUC AUC UCC UAC GGA 131 GIU Leu Asp Thr Tyr Thr MET Ile Ser Tyr Gly	6690 CCA GCU GAC ANA AGC GAC ANA GGU UUU GUU QUU GGU Pro Ala Asp Lys Ser Asp Lys Gly Phe Val Leu Gly	6810 NG ACC CUU GAG GCU AUC CUC UCC CCA CCC CCU JYS Thr Leu Clu Ala Ile Leu Ser Phe Ala Arg Arg	6930 NUC CAG CAC CUC UNU CAG AUC CCA ACC UAC ACA UCA The CIN CIY Leu Phe Ciu Ile Pro Ser Tyr Arg Ser	rac gec gua caa gug aaa age uco aaa gag quu uuc		and the corresponding amino acid viral nuctains are indicated and	ix referring to the following	
AUC GUC	GUU CAU Val Asp	Pro 000 Pro 000	Arg Cac Arg Lev	CAG CCA	AAC GAC Lys Asp	CAC UCA His Ser	uuu cau Phe Asp	GCC UAL Ala Tyr	GAG GG/ GIU GIY	AUC ACL Ile Thr	GCC UCA Ala Ser	GAG CCC	89		A ₁₀ 6	super	
Place	PSG Da	Asp	Ser	Ęg	Eec	MET	occ Ala	CAC	Tyr	Acc	MET	Phe	30		8	2 cc	
ACS Thr	Lec35	5715 Lys	5835 5835	5855 Ala	6075 UUC Phe	6195 CAA	6315 Ser	6435 61u	6555 CAC His	6675 CAA	6735 Val	6915 Leu	7035		, tyl	vith	
Asp Asp	[83	ASC	Tyr	D L C	AC	Ala	TYL	8 Brd Brd	ACA	25	P P C C	Argu Argu	ğ		Į Į		
2 V V V V	S C K C K C	n Sei	a CA(ប៉ុន្តិ៍	2 C C C C	ပ်ခို ပြော	P C A	PC AS	A V V V V	อีรี	Υ. Έ.Υ	9 <u>₹</u>	c nî			cate	
95 95	S. T. C.	n CC a CC	a CC	Ę2 در	20	3å Bå	20	n CC	re E	s Sei	e LAN D	ά. Ω	С С			ipu	
S C C	5 £C 9R2	FS BSS BSS BSS BSS BSS BSS BSS BSS BSS B	800 800	58	800 860	222	50 50 50	25 2,33	200	é Čô	35	55	22		gio	le le	
ΣΣξέ	Pro55	2222	222 222	12 Sel	28° 295⊊		20°	3 40 11 11	55 ⁶⁵⁴	666 P	562	9058 800	AC A4		j re	8	
DU G	55 25	200	280	se Se	AC V	ETC	AC CI	er A er A	202	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	P A F A F	LA DO	AU G.		ding	, dat	f
	SP C	56 00	he A D	SP T	AA T		CGA A	LC LC	yr G	VS C	yr o	99 99	GAG		C C	3) JCe	1
	al A	NG D	SC P	85	L & O	A R A	AC A	Sc.	Le U	NC A eu L	SPU T	AC U	۲		the	int_	•
CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA	Eec Sec	figer figer	200 P	5251	Lys L	Thr L	CAA CAA		ASh J	393 a		855 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	SS SS		of	- Se -	4
569 D	Ala 1	ASD 5	AM	MET 0.5	C.MC 6	Tyr	ALa	Ala Co	AAC	GAG 6	CAC 6				nce D	acid	1
S, CC	AAG	Phe	GAC	Ala	MET	Eeg	Phe	Ash	Leu	Phe	Phe	Lec	Ala		due	2	f
Tyr	MET	CUC	GAG	GAC	Eec	AUU	CAU His	Pro CC	AUU	GAC	CAC	CLY CLY	Asp		se	i mir	5
ទទួ	ANG	61y 61y	GAA	Eg	ANG	CAC	AcA	CAC	ACA	Eeg	ACA	ALa	0.00		tide		5
A a con	5550 GLN	5670 Tyr	Thr Thr CC	665 61 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	E E E E E S S	GAA GAA	500	2 Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second	S510 AAC Asn	Asp GAU Asp	AAA Lys	CUC CUC Val	258		leo	orti	נפ
A 80	Eg	Ala	S S S S S S S S S S S S S S S S S S S	ASP	- SC	Cou	Page 2	585	, NUC	Tyr	Leg_	ັບ ຮັດ	CUC	AAA	onu		5
AC	3Ē	CUU	AG	CUU	60 Ma	ទ្ធដ	ACA Arg	Phe	AUC	GAU Asp	UUC Phe	AUC	GCC	AAA	ete i	a sur	Ď
	Ser	PPC AC	C CCA	500	CAC	Leu	GIN	GAC	Ser	Ser	Tr Tr C	Leu	AAC	AAA	nple	1 2 2	2
s ACC	ທູບ ຊີຊີ	7 C C C C C	A CAC	e Lys	A GUL	5 CUU	- ICC	The AC	ACA	Ala) CUC	ANG	CUC Val	AMA	ទីទី	, the '	۵ 1
r Lyn	1022 Sec. 1	u 255	522 s	a 198	000	C 613	U CAL	چ چ	50°		c cal		1.000 3. Tr	700f	Ĕ	i i	+
2 L	SC GU	ದಿಸ ಲ್ಲಿ	S Ly	N ALC	ς Υ.Υ.	e va	P CCU	ЭĘ ЭĘ	s Sel	c cut e vai	P A C A C A	5 E	97 78	n CM	2 10 10	- × -	
5ĕ ₽E	೨೦ ಶ ಕಿ	רב.> זנ⊳	AA C	AC 00	20	SC AU	58 80	D S S S S S	89 25	NG NG	81 1	S P S C	5r 83	2	Ш		۲ ñ
2 S S S S S S S	UCC N CVS S	AAA A Lys Ti	Ser L	AUC U. Ile Ty	ACC Thr VS	ACU Thr A	AAC CC ASn Pr	CAC CAC CAC CAC	Ser	GAC C/ Asp As	CAG CAG CIN Se	CCC CCC SIV	cuu Ur Leu Ty	ច អូ		those	
																-	

Nucleotide Sequence

The nucleotide sequence determined from the recombinant clones and by primer extension sequencing is shown in Figure 2. The sequence includes an open translational reading frame of 2,333 codons terminating at a UAA codon 93 bases from the 3' poly (A) sequence. Two lines of evidence suggest that this represents the sequence coding for the FMDV polyprotein. Firstly, no other reading frame of significant length is found in the sequence. Secondly, protein sequence data is available for several viral proteins (11 and refs. therein, 14) and these sequences correlate with the major open reading frame. Further sequence towards the 5' end from the major open reading frame has termination codons in all three reading frames (15).

Examination of our sequence gels allows us to estimate the poly (C) tract to be at least 500 nucleotides from the initiation site for translation. This would mean that FMDV RNA has a 5' untranslated region in excess of 1,000 nucleotides long and a total genome length of at least 8,100 nucleotides.

The nucleotide composition and dinucleotide frequency of the coding region are consistent with the data previously obtained for the region of the genome coding for the structural proteins (11). In particular the base composition shows a bias in favour of C+G over A+U. This is also reflected in chemical analysis of the RNA (16). The relatively high frequency of CpG is also maintained throughout the genome being 83% of the expected frequency compared to 37% for eukaryotes in general (17) and 47% for poliovirus (18).

Predicted Amino Acid Sequence

The processing of picornaviral proteins is a complex process and appears highly variable between different members of the family. However, the general genomic organisation and the primary cleavage products derived from the polyprotein are similar, (1, 19) with one important exception. The region to the 5' side of the genome coding for the structural genes in FMDV and encephalomyocarditis virus code for additional proteins which have no identifiable counterparts in poliovirus. It has been shown that two proteins are coded for in this part of the genome in FMDV; namely p20a (Lab) and p16 (Lb) (for polypeptide nomenclature see Figure 1). Analysis of these proteins has shown them to have similar tryptic peptide maps and limited proteolysis patterns (6, 15). Both of these proteins can be labelled <u>in vitro</u> using N-formyl (35 -S) methionine t RNA_f (6) and studies involving a protease inhibitor suggest that the C-termini of these proteins are the same (20). These data suggest that there are two initiation sites in FMDV RNA. Examination of the sequence data presented here shows the presence of a second AUG codon located at nucleotide position 85 and in the same reading

frame as the initial AUG at position 1 (see Figure 2). Initiation at these two sites would result in two similar proteins with a molecular weight difference of 3,800, i.e. the difference in molecular weight between p20a (Lab) and p16 (Lb). Beck <u>et al</u> (21) have also recently reported the presence of these two putative initiation sites in two other serotypes of the virus.

Although no protein sequence data are available to confirm the predicted sequence of the three VPg proteins, their charge, amino acid composition and tryptic peptide composition (2) are entirely consistent with the sequences presented here. These data also agree with those of Forss and Schaller (22) on the predicted sequence of the VPgs from FMDV serotypes O and C. Although in general FMDV and polio show a high degree of similarity in genome organisation the VPg genes are an example where they differ significantly. Polio has a single VPg gene whereas FMDV has three, which, moreover are highly conserved in three serotypes (22). All three VPgs from FMDV are equally represented in RNA extracted from virus particles (2).

The poliovirus VPg has been found in a precursor molecule of molecular weight 12,000 of which the VPg protein itself is contained in the C-terminal portion. The non-VPg sequences in this precursor contain a hydrophobic region of 22 amino acids situated 7 amino acids to the N-terminal side of VPg. This hydrophobic area is thought to be responsible for the association of the VPg precursor with cellular membranes, since the VPg precursor is found in membrane bound complexes (23). Comparison of hydrophobic region in FMDV comparable to that in polio. The functional precursor, (if any), containing the FMDV VPgs is at present unknown.

Comparison of our sequence with the recently published sequence of the FMDV A_{12} polymerase (p56a; 3d) (14) show them to be very similar. There are 16 amino acid changes (out of 470 total) which are distributed throughout the protein. <u>Cleavage sites</u>

The proteolytic cleavage sites involved in the processing of the poliovirus polyprotein into the final products show a high degree of uniformity not seen with FMDV. In polio eight of the identified cleavage sites are between gln-gly pairs; of the remaining cleavages one is between tyr-gly and one is between asn-ser (24). The latter cleavage is between VP4 and VP2 and occurs during the final stages of virus maturation. With FMDV, on the other hand, little homology is apparent in the sequences around the known cleavage sites. The possible cleavage sites involved in processing the polyprotein of FMDV are indicated in Figure 2. Those sites for which some amino acid sequence data are available are also indicated. Both host

and viral specified proteases are involved in the processing of FMDV proteins (20).

The cleavages thought to be caused by a viral enzyme (secondary cleavages) indicate a preference for glu-gly bonds, for example between VP2 and VP3, at the N termini of the VPgs and between p20b (3c) and p56a (3d, polymerase). Of the twelve putative cleavage sites (both primary and secondary) five occur between glutamate and glycine.

The putative host specified cleavage site (primary cleavage) between p20a/p16 and p88 (L and P1) has been revised from previous reports (7). This cleavage is now proposed to occur after amino acid 204 between gly and gln on the basis of homology between FMDV and EMC VP4 proteins (A. Palmenberg, personal communication), and our recent observation that FMDV VP4 contains proline (15). With the revised cleavage site VP4 contains a single proline at amino acid position 208 (see Figure 2). The N-terminus of VP4 is known to be refractory to Edman degradation in all picornaviruses which have been examined (11, 19, 24, 25) and with FMDV and EMC this blockage may be due to deamination and cyclisation of a glutamine.

CONCLUDING REMARKS

At present we do not have clones which span the poly (C) tract and so it has not been possible to analyse the sequence of the 5' non coding region of the FMDV genome. The reason for the lack of 5' end clones is not clear. It could be simply due to the low frequency of cDNA transcripts which extend to the extreme 5' end, or alternatively the unusual structure of the poly (C) tract may be unstable in <u>E.</u> <u>coli</u> and eliminated during plasmid replication. Direct RNA sequencing of the, extreme 5' end of the genome has been reported (26) and shows that considerable homology exists between the first 27 nucleotides from all seven serotypes of FMDV. This suggests that this region has a critical role in virus replication.

Sequence data from the 3' untranslated region of FMDV have also been reported (14, 27). Porter <u>et al</u> (27) analysed the nucleotide sequence adjacent to the poly (A) for five serotypes and found ~75% homology in the first 20 nucleotides and a highly conserved sequence of eleven nucleotides at position -7 to -17 from the poly (A) tract. One of these serotypes was identical to the virus used in this study and the sequence of 42 nucleotides reported agrees exactly with that reported here. Recently the 3' sequence of a second type A virus (A₁₂) has been reported (14) and maintains the sequence of the eleven highly conserved nucleotides, however the remainder of the sequence up to the poly (A) is considerably different and includes three additional nucleotides. In the A₁₂ virus there is a single UAA termination codon located 96 nucleotides from the poly (A) tract in contrast to the single UAA termination codon located 93 nucleotides from the poly (A) reported here from the $\rm A_{10}$ virus.

A knowledge of the coding sequence of a virus is a necessary pre-requisite for the full understanding of the functions of the encoded proteins in virus structure and replication. Moreover, comparisons of the complete sequences of different viruses from within the picornavirus family will indicate evolutionary relationships between the virus groups. Finally the sequence data is essential for experiments involving site specific mutagenesis of an infectious clone of virus RNA (28) in order to modify specific regions of the genome.

ACKNOWLEDGEMENTS

We would like to thank Dr A Makoff for computer analysis and Dr F Brown for helpful discussion and critical reading of the manuscript. We would like to thank Dr M Eaton (Cell Tech Ltd, UK) for kindly providing the synthetic oligonucleotide primer.

*Present address: Wellcome Biotechnology Limited, Ash Road, Pirbright, Woking, Surrey, UK

REFERENCES

- 1. Sangar DV. (1979) J. Gen. Virol. 45, 1-13.
- 2. King AMQ, Sangar DV, Harris TJR and Brown F. (1980) J. Virol. 34, 627-634.
- 3. Sangar DV, Rowlands DJ, Harris TJR and Brown F. (1977) Nature 268, 648-650.
- 4. Flanegan JB, Petersson RF, Ambros V, Hewlett MJ and Baltimore D. (1977) PNAS USA 74, 961-965.
- 5. Nomoto A, Detjen B, Pozzatti R and Wimmer E. (1977) Nature 268, 208-213.
- 6. Sangar DV, Black DN, Rowlands DJ, Harris TJR and Brown F. (1980) J. Virol. 33, 59-68.
- Boothroyd JC, Highfield PE, Cross GAM, Rowlands DJ, Lowe PA, Brown F and Harris TJR. (1981) Nature 290, 800-802.
- 8. Wensink PC, Finnegan DJ, Donelson JE and Hogness DS. (1974) Cell 3, 315-325.
- 9. Woods DE, Crampton JM, Clarke BE and Williamson R. (1980) Nucleic Acids Research 8, 5157-5168.
- Maxam A and Gilbert W. (1980) in Methods in Enzymology Vol. 65, pp499-560, Grossman L and Moldave K (Eds), Academic Press, N.Y.
- 11. Boothroyd JC, Harris TJR, Rowlands DJ and Lowe PA. (1982) Gene 17, 153-161.
- 12. Rowlands DJ, Clarke BE, Carroll AR, Brown F, Nicholson BH, Bittle JL, Houghten RA and Lerner RA. (1983) Nature 306, 694-697.
- 13. Makoff AJ, Paynter CA, Rowlands DJ and Boothroyd JC. (1982) Nucleic Acids Research 10, 8285-8295.
- 14. Robertson BH, Morgan DO, Moore DM, Grubman MJ, Card J, Fischer T, Weddell G, Dowbenko D and Yansura D. (1983) Virology 126, 614-623.
- 15. Clarke BE (1984) Manuscript in preparation.
- 16. Newman JFE, Rowlands DJ and Brown F. (1973) J. Gen. Virol. 18, 171-180.
- 17. Nussinov R. (1981) J. Mol. Biol. 149, 125-131.
- 18. Racaniello VR and Baltimore D. (1981) PNAS USA 78, 4887-4891.
- 19. Rueckert RR. (1976) in Comprehensive Virology, Fraenkel-Conrat H and

Wagner RR (Eds) Vol. 6 pp131-213, Plenum Press, New York.

- 20. Burroughs JN, Sangar DV, Clarke BE and Rowlands DJ. (1984) J. Virol. In press.
- 21. Beck E, Forss S, Strebel K, Cattaneo R and Feil G. (1983) Nucleic Acids Research 11, 7873-7885.
- 22. Forss S and Schaller H. (1982) Nucleic Acids Research 10, 6441-6450.
- 23. Semler BL, Anderson CW, Kitamura N, Rotherberg PG, Wishart WL and Wimmer E. (1981) PNAS USA 78, 3463-3468.
- Kitamura N, Semler BL, Rothberg PG, Larsen GR, Adler CJ, Dorner AJ, Emini EA, Hanecak R, Lee JJ, van der Werf S, Anderson CW and Wimmer E. (1981) Nature 291, 547-553.
- Scraba DG. (1979) in The Molecular Biology of Picornaviruses, Perez-Bercoff R. (Ed) pp1-23, Plenum Press, New York.
- 26. Harris TJR. (1980) J. Virol. 36, 659-664.
- 27. Porter AG, Fellner P, Black DN, Rowlands DJ, Harris TJR and Brown F. (1978) Nature 276, 298-301.
- 28. Racaniello VR and Baltimore D. (1981) Science 214, 916-919.