The EMBO Journal Vol.2 No.4 pp.555-559, 1983

The molecular basis of the antigenic variation of foot-and-mouth disease virus

E.Beck*, G.Feil and K.Strohmaier[1]

Department of Microbiology, University of Heidelberg, Im Neuenheimer Feld 230, D-6900 Heidelberg, and [1]Federal Research Institute for Animal Virus Diseases, Paul-Ehrlich-Str. 28, D-7400 Tübingen, FRG

Communicated by H.Schaller
Received on 17 January 1983; revised on 10 February 1983

We have cloned and sequenced the viral protein (VP1)-coding regions of two foot-and-mouth disease virus (FDMV) serotypes (C_1 and A_5). Comparison of the derived amino acid sequences with the known VP1 sequence of FMDV O_1K and the two FMDV A subtypes A_{10} and A_{12} shows two highly variable regions in the protein, at positions 40-60 and 130-160, as possible antigenic sites. In both variable regions, several sites could be detected where all three sequences of the A subtypes are identical but the three types A, C and O differ from each other. The second variable region overlaps with a major immunogenic determinant of the virus.

Key words: FMDV/amino acid sequence/immunogenic sites/antigenic variation

Introduction

Foot-and-mouth disease virus (FMDV) is able to escape the host immune system by a high variability of its surface antigens. Seven distinct serotypes and >60 subtypes of the virus have been characterized by serological analysis. The major antigenic determinants are located in a single viral protein, VP1 (Bachrach *et al.*, 1975), and potentially immunogenic regions have been identified in the C-terminal third of this protein (Strohmaier *et al.*, 1982). Comparison of the VP1 structure between serotypes on the one hand and between subtypes of a certain serotype on the other hand should allow us to localize potentially immunogenic sites and to distinguish between alterations which lead to a change in antigenicity and alterations which do not.

A complete amino acid sequence of VP1 was derived from the nucleotide sequence of cloned cDNA of serotype O (Kurz *et al.*, 1981). We have now obtained the amino acid sequence of this protein from cloned FMDV cDNAs of two other European serotypes, A and C. Corresponding sequences of two additional FMDV A subtypes were also determined in other laboratories (FMDV A_{12}: Kleid *et al.*, 1981, FMDV A_{10}: Boothroyd *et al.*, 1982).

A comparison of the VP1 primary structures of the different FMDV strains described here shows two variable regions. Combined with immunological data from enzymatically or chemically derived cleavage products of VP1 (Strohmaier *et al.*, 1982) and computer predicted secondary structures of this protein, these results lead to the synthesis of a potential peptide vaccine against FMDV (Pfaff *et al.*, 1982).

Results

Cloning of the coat protein-coding region

Identification and isolation of VP1 encoding cDNA clones from FMDV A5 and strain C_1O was facilitated by the availability of corresponding clones from strain O_1K (Küpper *et al.*, 1981). Preliminary cross-hybridisation experiments with these clones had indicated that the VP1 encoding segment of the FMDV RNA was highly variable between serotypes whereas the adjacent region coding for non-structural proteins was not (Kurz *et al.*, 1981). To identify more precisely the homologous region next to the VP1 gene, a series of short, segment-specific probes was prepared by *Hae*III digestion of the appropriate single-stranded FMDV cDNA clone (see Materials and methods) and tested for stable annealing to the RNA of the heterologous serotypes. There was no cross-hybridisation with probes from within the O_1K VP1 gene (positions 2892-3522 and 3523-3542), whereas probes from the adjacent cDNA segment cross-hybridised (positions 3662-3752, 3753-3792, 3793-3882, 3883-3921, 3932-4017). Therefore, these latter DNA fragments could be used to identify clones from the VP1 region or to prime specifically cDNA synthesis on heterologous FMDV RNA next to the VP1 gene.

The latter method was used with strain C_1O. A short single-stranded *Hae*III fragment from O_1K cDNA covering FMDV RNA from position 3793 to 3882 was used to prime cDNA synthesis on C_1 RNA. The single-stranded cDNA was rendered double-stranded by *Escherichia coli* DNA polymerase I as described (Küpper *et al.*, 1981) and fitted with synthetic *Eco*RI linkers. As the primer fragment covered a *Hind*III site ~230 bp downstream from the end of VP1, the cDNA was cleaved with *Eco*RI and *Hind*III, sized by gel electrophoresis and fragments >1000 nucleotides were ligated into the vector pBR322 *via* the *Eco*RI and *Hind*III sites. The largest clone (pFMDVC1-9) recovered using this approach contained a 3000-bp cDNA fragment extending from the *Hind*III site at position 3850 up to position 900, thus entirely covering the region coding for the viral coat proteins VP1-VP4 (see Figure 1).

The cDNA clone from FMDV strain A_5 used in this analysis, pFMDVA5-1, was a gift from H.Küpper. It was constructed in the same way as the cDNA clones from strain O_1K by oligo(dT) priming of the polyadenylated viral RNA (H.Küpper *et al.*, 1981) and identified by screening with a labelled DNA fragment of pFMDV1034 from the conserved region adjacent to VP1. As determined by the DNA sequence analysis described here and as shown in Figure 1, clone pFMDVA5-1 covers the complete VP1 region and some additional 1300 bp downstream. To simplify its sequence analysis, the FMDV insert of the plasmid was subcloned, after cleavage with *Pst*I and *Hind*III, into vector pUC9 (Vieira and Messing, 1982). The plasmid containing the VP1 gene was designated pFMDVA5-1a.

Nucleotide sequence of the VP1 region of FMDV C_1 and FMDV A_5

Nucleotide sequences from pFMDVA5-1a and pFMDVC1-9 were determined according to Maxam and Gilbert (1980) as outlined in the lower part of Figure 1. Both cloned DNA fragments were sequenced completely. Only the parts of the sequences encoding VP1 are shown in Figure 2.

*To whom reprint requests should be sent.

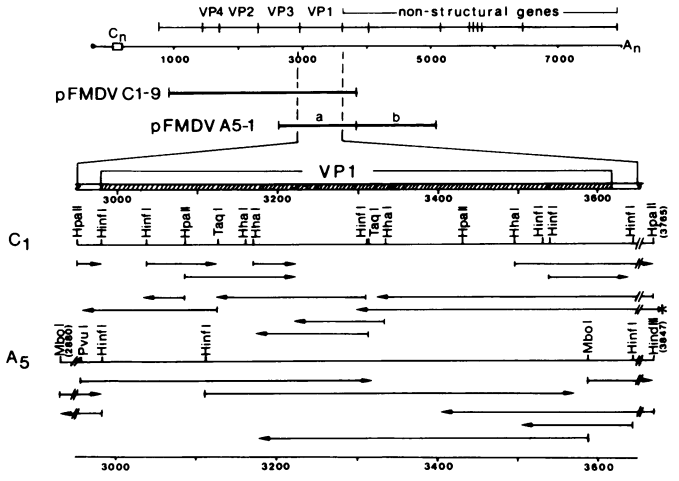

Fig. 1. Location of the cDNA clones pFMDVC1-9 and pFMDVA5-1 in the genome of FMDV and sequencing strategy applied. For numbering of the FMDV RNA the system of Küpper *et al.* (1981) was used. C_n and A_n refer to the internal poly(C) tract and the polyadenylated 3′ end. The VP1 encoding region is enlarged in the lower part of the figures. Restriction endonuclease maps for enzymes used in the sequence analysis of clones pFMDVC1-9 and pFMDVA5-1a are shown. Direction and extent of the individual sequencing runs are indicated by horizontal arrows. The sequence obtained by primer extension with reverse transcriptase on C_1 RNA (see text) is indicated by an asterisk.

Nucleotide sequences from the C-terminal part of the VP1 gene were also directly determined from the viral RNAs by primer extension with reverse transcriptase. A small single-stranded *Hae*III fragment from O_1K (position 3662–3752) was ^{32}P end-labelled, annealed onto FMDV RNA, elongated by the reverse transcriptase reaction and the resulting cDNA copies were used without further purification for sequence analysis according to Maxam and Gilbert (1980). The sequences obtained (not shown) were in complete agreement with the results obtained from the cDNA clones as described above. As cDNA sequences can be read over all of the antigenic determinant in the C-terminal part of the VP1 gene (see below) this direct sequencing method may be useful for rapid analysis of sequence variations in field strains of FMDV.

To determine the borders of the VP1 gene, the nucleotide sequences from serotypes C_1 and A_5 were aligned to the O_1K sequence (Figure 2) for which the gene limits have been identified using amino acid sequence data from the gene product (Strohmaier *et al.*, 1978; Kurz *et al.*, 1981). Although there are stretches of high variability between the nucleotide sequences of the three different serotypes, the regions encoding the termini of VP1 were sufficiently conserved to allow an unambiguous alignment. From the homology to the O_1K nucleotide sequence, we conclude that in FMDV C_1 and FMDV A_5 the VP1 coding sequences start with nucleotide 2977 and stop at positions 3603 and 3612 in C_1 and A_5,

Fig. 2. Comparison of the VP1 encoding nucleotide sequences of FMDV serotypes C_1, O_1K and A_5. The sequences are aligned according to homologies in the more constant regions. The somewhat arbitrary alignment around position 3400 corresponds to the amino acid alignment in Figure 3. The numbering refers to the O_1K nucleotide sequence (Kurz *et al.*, 1981). Strokes represent deletions used to optimize alignment and asterisks indicate identical nucleotides in all three serotypes.

respectively. Hence, it follows that the length of the VP1-coding region varies slightly between the different serotypes of FMDV: the O_1K nucleotide sequence codes for a protein of 213 amino acid residues and the sequences of types C_1 and A_5 code for 209 and 212 amino acids, respectively.

Comparison of the C_1 and A_5 nucleotide sequences with each other and with the O_1K sequence shows a similar degree of homology between the three serotypes within the VP1 encoding region: C_1/A_5 58%, C_1/O_1K 64% and A_5/O_1K 65% homology, respectively. 52% of the nucleotides are common in all three serotypes. Homology of the A_5 nucleotide sequence with the corresponding sequences of the subtypes A_{10} (Boothroyd *et al.*, 1982) and A_{12} (Kleid *et al.*, 1981) is expectedly higher: A_5/A_{12} 90%, A_5/A_{10} 87%, A_{12}/A_{10} 91%).

Base changes between serotypes as well as between subtypes are clustered mainly in two regions (positions 3100–3160 and 3370–3450; see Figure 2). Changes outside these two regions often concern third positions in codons and do not lead to amino acid exchanges. Determination of the codon usage in C_1 and A_5 showed a similar bias in favour of C and G over A and T in the third position of codons (57% and 66% for C_1 and A_5, respectively) as was observed before with O_1K and A_{10} (Kurz *et al.*, 1981; Boothroyd *et al.*, 1982).

There is a striking homology of the nucleotide sequences beyond the VP1 encoding region of ~90% between all three serotypes. This homology extends from position 3620 to the *Hind*III site at position 3850 to which all three sequences have so far been determined (not shown). This agrees with the results of the cross-hybridisation experiments described above.

Amino acid exchanges in VP1 between different serotypes of FMDV are clustered in two regions

The amino acid sequences of the VP1 region are compiled in Figure 3. They are derived from the nucleotide sequences of the three serotypes C_1, A_5 and O_1 and two additional FMDV A subtypes A_{12} (Kleid *et al.*, 1981) and A_{10} (Boothroyd *et al.*, 1982). This comparison shows that some 60% of the amino acid residues are conserved between all three serotypes. The most conserved regions are the N-terminal portion up to position 40, the middle region (position 60–130), and the C-terminal quarter of the protein (from position 160). These more constant parts are interrupted by two variable regions, one between position 40 and 60, the other between position 130 and 160. In the former region, 16 out of 20 amino acid residues are exchanged. Nine of them represent variation between all three, the remaining seven between only two serotypes. In general, these changes do not affect the charge distribution of this part of the protein. The second variable region is more extended. It covers ~30 amino acid residues of which about two thirds are exchanged. The majority of the exchanges represent variation between all three serotypes. In contrast to the first variable region, many of the exchanges involve charged amino acids and therefore the distribution of acidic and basic amino acid residues in this region differs between all three serotypes. Around position 140 there is a hypervariable region where the amino acid sequences also differ in length by up to four amino acid residues.

As the three FMDV A subtypes are antigenically closely related, their immunogenic sites should also contain identical amino acids in regions where the three serotypes differ from each other. Such positions (boxed in Figure 3) are clustered in a few regions: positions 44–60 (7x), 81–85 (3x), 96, 134–137 (3x), 149–158 (4x) and 199.

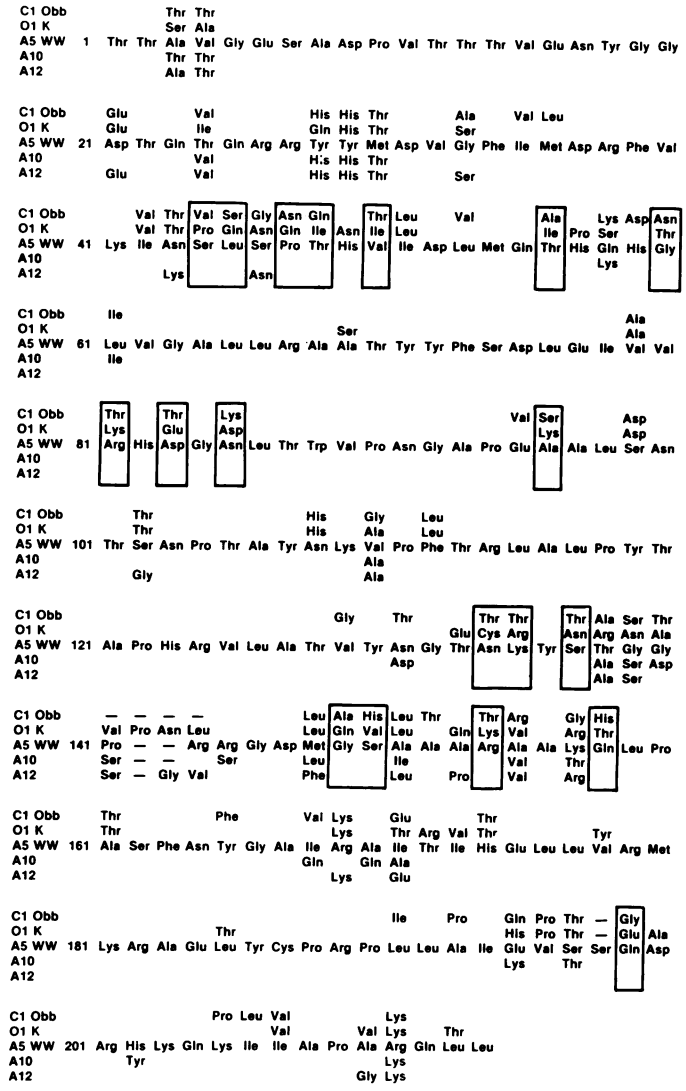

Fig. 3. Comparison of the VP1 amino acid sequences of FMDV strains C_1, O_1K, A_5, A_{10} and A_{12}. The A_5 sequence is shown completely. For the remaining strains, only differences to the A_5 sequence are specified. Strokes represent deletions used to optimize alignment of the sequences. Positions which are different in all three serotypes (upper 3 lines) but identical in the A subtypes (lower 3 lines) are boxed.

Discussion

FMDV seems to us to be a suitable model system to study the molecular basis of the immunogenicity of a viral capsid protein for two reasons. The major antigenic determinants of the virus seem to be localized on a single viral protein, VP1 (Bachrach *et al.*, 1975) and there are many serologically different strains of the virus available. The knowledge of the structural features of the immunogenic sites in FMDV is also important since it should be possible by recombinant DNA technology or by chemical synthesis to simulate these sites and to construct synthetic vaccines (Bittle *et al.*, 1982; Pfaff *et al.*, 1982).

A comparison of the VP1 primary structure of three serotypes and three subtypes of FMDV shows that there are essentially two highly variable regions within the protein, one between positions 40 and 60 and another between positions

130 and 160. Other parts of the amino acid sequence are highly conserved (e.g., positions 5–20, 61–80, 113–128) and the genetic drift in the corresponding coding sequence is almost exclusively confined to silent third base mutations in the viral RNA (see Figure 2). Thus, these amino acid sequences must be important for the functional integrity of the VP1 subunit in FMDV. Significant restrictions to amino acid changes also seem to exist for the first variable region, since changes stay generally within the same functional groups of amino acids. Furthermore, only few differences are found in this VP1 segment between the evolutionary closely related subtypes from serotype A. In contrast, a high sequence variation even between subtypes is present in the second variable region.

Because of the high antigenic diversity in FMDV, potential antigenic determinants are expected to overlap at least in part with variable regions in the VP1 sequence and also to contain serotypically characteristic positions, where amino acids are common between subtypes but different between serotypes. Furthermore, as hydrophilic surface components of the virion, antigenic determinants are also expected to contain polar residues in their amino acid sequence.

From the presently available sequence data, the second variable region appears to meet best these requirements for an antigenic determinant and was therefore expected to cover a major immunogenic site of FMDV (Pfaff *et al.*, 1982). In addition, in FMDV O_1K, part of the second variable region (positions 144–159) contains a biased distribution of polar and apolar residues when projected in a computer predicted α-helix (Pfaff *et al.*, 1982). This one-sided stable location of polar residues can also be predicted for the other strains in the corresponding region, although in A_{12} a proline residue within the sequence interrupts an ideal helical structure. Complete virions incubated with proteases are cleaved at this site in the protein, confirming that this region is exposed at the surface of the virus (Strohmaier *et al.*, 1982).

Proof that the predicted α-helical part of the second variable region is in fact an antigenic determining site comes from immunological data with antisera against proteolytically or chemically derived fragments of VP1 (Strohmaier *et al.*, 1982) and with synthetic oligopeptides of serotype O_1K which were successfully used to induce virus neutralising antibodies. Interestingly, the most exchanged part of the variable region (positions 137–144) seems not to be necessary for serospecificity since it is not contained in the highly serospecific peptides, positions 144–159 (Pfaff *et al.*, 1982), and positions 141–160 (Bittle *et al.*, 1982). The most conclusive evidence that these peptides cover the dominant immunogenic site of FMDV comes from the fact that a carrier-coupled peptide bound virtually all (99%) of the anti-FMDV antibodies in serum (Pfaff *et al.*, 1982). Peptides corresponding to the sequence close to the C terminus of VP1 also evoke virus-specific antibodies (Strohmaier *et al.*, 1982; Bittle *et al.*, 1982). However, their virus neutralising activity lies below the values of sera against the oligopeptides of the second variable region.

The highest virus neutralising activities reached in antisera against peptides lie several orders of magnitude below the immunogenic efficiency of complete virions. Therefore, the ability of this amino acid sequence to determine serospecificity and to bind essentially all of the virus neutralising antibodies may reflect that it represents the 'core' of the dominant immunogenic site of FMDV. In the virus the structure of the antigenic determinant must be more complex and probably depends on a particular tertiary struture of VP1. This structure seems to be maintained correctly in the intact virion only, but not in isolated VP1 or in the 12S subunit of the virus coat protein containing VP1, VP2 and VP3 (Cartwright *et al.*, 1982).

Materials and methods

Enzymes

Restriction endonucleases (*Eco*RI, *Hind*III, *Pst*I, *Hae*III, *Hinf*I, *Hha*I, *Hpa*II and *Taq*I) were purified essentially as described by Roberts *et al.* (1976), *Pvu*I and *Mbo*I were purchased from Boehringer Mannheim and New England Biolabs, respectively. T4 polynucleotide kinase was purified by a modification of the procedure of Richardson (1965), calf intestinal phosphatase and S1 exonuclease were purchased from Boehringer Mannheim and New England Biolabs, respectively.

Single-stranded FMDV cDNA clones

The FMDV part of plasmid pFMDV1034 (Küpper *et al.*, 1981) was inserted *via* the *Pst*I sites into the single-stranded DNA vector fd109 (Herrmann *et al.*, 1980) in orientation opposite to the viral RNA. To obtain another single-stranded DNA clone with a less complex *Hae*III pattern, a smaller FMDV fragment coding for the C-terminal half of VP1 plus some 230 adjacent nucleotides was inserted in the fd109 derivative fd109.2 (R.Herrmann, as cited in Weiher and Schaller, 1982) as follows: an 800-bp *Hha*I fragment from plasmid pFMDV1034 (Kurz *et al.*, 1981) containing an internal *Hind*III site (position 3847 in the O_1K nucleotide sequence) was isolated, incubated with DNA polymerase I in the presence of the four dNTPs to digest the protruding 3′ ends and joined with synthetic *Eco*RI linkers. After cleavage with *Eco*RI and *Hind*III, the larger of the two resulting fragments (513 bp) was inserted in fd109.2 *via* the corresponding restriction sites.

Hybridisation experiments

Single-stranded DNA isolated from phages was digested with *Hae*III, the resulting fragments 5′ end-labelled with polynucleotide kinase and [γ-^{32}P]-ATP and separated electrophoretically. 0.1 pmol each of the end-labelled *Hae*III fragments from within or from outside of the VP1 coding region were incubated with an equimolar amount and with a 3-fold excess of viral RNA of serotypes O_1K, A_5 and C_1 in 10 μl 0.2 M NaCl, 50 mM Tris pH 8.0 for 30 min at 65°C. The mixtures were then separated electrophoretically on 6% polyacrylamide gels and the radioactivity determined at and above the initial positions of the fragments.

cDNA synthesis

2 pmol FMDV RNA were hybridized with 2 pmol of an end-labelled, single-stranded *Hae*III fragment in 5 μl 0.2 M KCl, 50 mM Tris pH 8.0 for 10 min at 65°C, then were added 60 μl of 50 mM Tris pH 8.3, 7 mM $MgCl_2$, 30 mM KCl, 1 mM dithioerythritol, dATP, dCTP, dGTP and dTTP each at 0.5 mM and 16 units of avian myeloblastosis virus reverse transcriptase and the mixture incubated for 1 h at 42°C. For the purpose of nucleotide sequence analysis, this cDNA was desalted by Sephadex G 150 gel filtration and subjected to the chemical method. To synthesize second strand cDNA, 2 μg RNase were added to 30 μl of the above mixture, the sample was incubated at 100°C for 3 min and for a further 15 min at 37°C. For second strand synthesis, 60 μl of 50 mM potassium phosphate buffer pH 7.0, 7 mM $MgCl_2$, 100 mM KCl, 0.1 mM each of dATP, dCTP, dGTP and dTTP and 6 units DNA polymerase I were added and the mixture incubated for 1 h at 15°C. To stop the reaction, the sample was extracted with phenol and passed through a Sephadex G-150 column. The double-stranded cDNA was treated with S1 nuclease and *Eco*RI linkers were ligated to the ends. As the primer fragment (positions 3792–3882 in the O_1K nucleotide sequence) carries a *Hind*III site, the DNA was cleaved with *Eco*RI and *Hind*III and sized on a 1% agarose gel. Fragments >1000 bp were eluted and inserted into plasmid pBR322 *via* the *Eco*RI and *Hind*III sites.

Nucleotide sequence analysis

DNA sequencing was performed essentially as described by Maxam and Gilbert (1980). Sequencing gels were dried in order to enhance band sharpness and to shorten exposure times (Garoff and Ansorge, 1981). Nucleotide sequences were stored and processed using the computer programs of Osterburg *et al.* (1982).

Acknowledgements

We thank H.Küpper for the gift of the clone pFMDVA5-1, H.Schaller for discussion and P.Shaw for help with the manuscript. Most of the C_1 nucleotide sequence was determined in a student course in the Winter term 1981/82. The work was supported by BIOGEN S.A. and the Deutsche Forschungsgemeinschaft, Forschergruppe Genexpression.

References

Bachrach,H.L., Moore,D.M., McKercher,P.D. and Polatnick,J. (1975) *J. Immunol.,* **115**, 1636-1641.

Bittle,J.L., Houghten,R.H., Alexander,H., Shinnick,T.M., Sutcliffe,J.G., Lerner,R.A., Rowlands,D.J. and Brown,F. (1982) *Nature,* **298**, 30-33.

Boothroyd,J.C., Harris,T.J.R., Rowlands,D.J. and Lowe,P.A. (1982) *Gene,* **17**, 153-161.

Cartwright,B., Morrell,D.J. and Brown,F. (1982) *J. Gen. Virol.,* **63**, 375-381.

Garoff,H. and Ansorge,W. (1981) *Anal. Biochem.,* **115**, 454-457.

Herrmann,R., Neugebauer,K., Pirkl,E., Zentgraf,H.W. and Schaller,H (1980) *Mol. Gen. Genet.,* **177**, 231-242.

Kleid,D.G., Yansura,D., Small,B., Dowbenko,D., Moore,D.M., Grubman, M.J., McKercher,P.D., Morgan,D.O., Robertson,B.H. and Bachrach, H.L. (1981) *Science (Wash.),* **214**, 1125-1129.

Küpper,H., Keller,W., Kurz,C., Forss,S., Schaller,H., Franze,R., Strohmaier,K., Marquardt,O., Zaslavsky,V.G. and Hofschneider,P.H. (1981) *Nature,* **289**, 555-559.

Kurz,C., Forss,S., Küpper,H., Strohmaier,K. and Schaller,H. (1981) *Nucleic Acids Res.,* **9**, 1919-1931.

Maxam,A.M. and Gilbert,W. (1980) in Grossmann,L. and Moldave,K. (eds.), *Methods in Enzymology,* Vol. **65**, Academic Press, NY, pp. 499-560.

Osterburg,G., Glatting,K.H. and Sommer,R. (1982) *Nucleic Acids Res.,* **10**, 207-216.

Pfaff,E., Mussgay,M., Böhm,H.O., Schulz,G.E. and Schaller,H. (1982) *EMBO J.,* **1**, 869-874.

Richardson,C.C. (1965) *Proc. Natl. Acad. Sci. USA,* **54**, 158-165.

Roberts,R.J., Myers,P.A., Morrison,A. and Murray,K. (1976) *J. Mol. Biol.,* **103**, 199-208.

Strohmaier,K., Wittmann-Liebold,B. and Geissler,A.W. (1978) *Biochem. Biophys. Res. Commun.,* **85**, 1640-1645.

Strohmaier,K., Franze,R. and Adam,K.H. (1982) *J. Gen. Virol.,* **59**, 295-306.

Vieira,J. and Messing,J. (1982) *Gene,* **19**, 252-268.

Weiher,H. and Schaller,H. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 1408-1412.