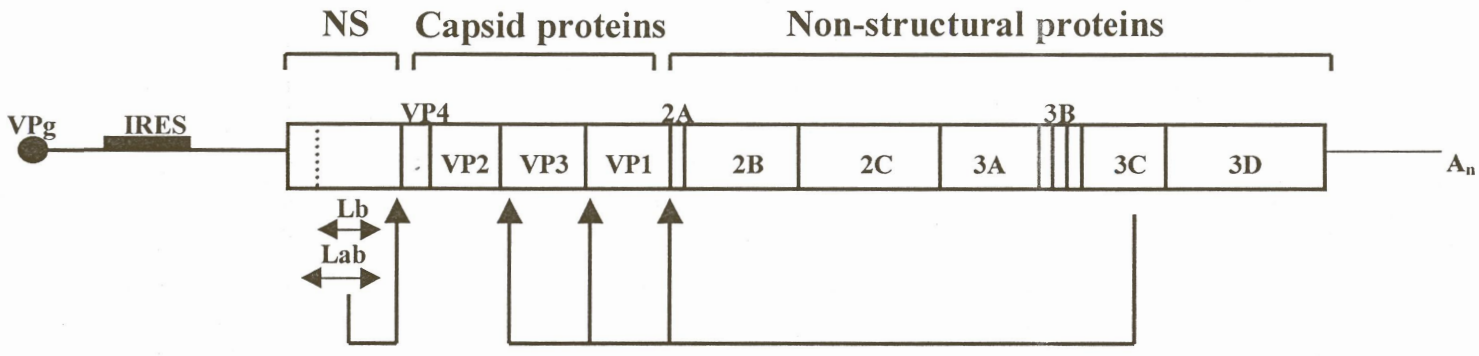
CHAPTER 3

Genetic heterogeneity in the foot-and-mouth disease virus leader and 3C proteinases

3.1 Introduction

The positive strand RNA genome of the foot-and-mouth disease virus is translated into a polyprotein which undergoes several proteolytic processing events to produce three polyprotein precursors P1, P2 and P3 (Domingo *et al.*, 1990; Belsham, 1993). The capsid proteins, which determine the antigenic characteristics of the virus, form part of the P1-2A polyprotein precursor (Figure 1.2, Chapter 1). The precursor is subsequently processed to produce the mature structural or capsid proteins, VP1, VP2, VP3 and VP4, which are assembled into virions (Figure 3.1). The precursor is obtained following cleavage at the L/P1 junction by the Leader proteinase (L^{pro}) (Strebel & Beck, 1986) as well as the 2A/2B junction by the 2A proteinase (Ryan *et al.*, 1991). The mechanism of the latter cleavage mediated by the 18 amino acid 2A region, is however not clear. The P1-2A precursor is then cleaved by the 3C proteinase ($3C^{pro}$) to produce VP0, VP3 and VP1. In addition to the viral protein processing, the $3C^{pro}$ of FMDV cleaves the host-cell protein histone H3 and may be involved with shut-off process of host-cell transcription (Falk *et al.*, 1990; Tesar & Marquardt, 1990).

The Leader proteinase acts both intra-and intermolecularly. Intramolecularly cleavage involves, as mentioned before, the cleavage of the L/P1 junction (RKLK↓GAGN) (Kirchweger *et al.*, 1994). During virus replication, L^{pro} specifically cleaves a host cell protein, eukaryotic initiation factor 4G (eIF-4G) (Devaney *et al.*, 1988). This event impairs the ability of the host cell to translate its own capped mRNAs. Characteristic of L^{pro} and conserved in all seven serotypes is the presence of two initiation sites (Sanger *et al.*, 1987). Both forms Lab and Lb, exhibit the processing functions (Medina *et al.*, 1993; Cao *et al.*, 1995). The latter function (eIF-4G processing) is



Structural / capsid proteins: **Non-structural proteins:**

VP4 = 1A **Leader = proteinase**

VP2 = 1B **2A = proteinase**

VP3 = 1C **3C = proteinase**

VP1 = 1D **3D = RNA - dependent RNA polymerase**

Figure 3.1: Proteolytic processing of the foot-and-mouth disease vial P1 region. Cleavage by the Leader (L/P1) and 3C proteinases (VP2/VP3; VP3/VP1; VP1/2A) are indicated.

executed in other picornaviruses (e.g. entero- and rhinovirus) by the 2A proteinase (Ryan & Flint, 1997). Different cleavage sites are used to cleave the eIF-4G by the Leader and 2A proteinases and are only seven amino acids apart (Kirchwegen *et al.*, 1994). Evidence for the cleavage of eIF-4G as well as eIF-4A by the 3C proteinase was recently obtained (Belsham *et al.*, 2000).

It was previously proposed, through amino acid sequence alignments, that the Leader proteinase of FMDV is a cysteine proteinase (Gorbalenya *et al.*, 1991). This hypothesis was confirmed by site-directed mutagenesis (Piccone *et al.*, 1995b; Roberts & Belsham, 1995) indicating that Cys^{51} and His^{148} are the active site residues. Recently, the crystal structure of the FMDV Leader proteinase from O_1Kaufbeuren was determined (Guarné *et al.*, 1998), confirming Cys^{51} and His^{148} to be the catalytic residues, situated on top of a deep cleft. Asp^{163} was observed to be the third member of the catalytic triad and is responsible for the correct orientation of His^{148} with respect to the nucleophilic cysteine. The structure of the FMDV Leader proteinase catalytic domain is reminiscent of that of the cysteine proteinase superfamily. The corresponding residues forming the catalytic triad in the cysteine proteinase, papain, are Cys^{25} – His^{159} – Asn^{175} (Ryan & Flint, 1997). It was also seen that the Leader proteinase contains a C-terminal extension, which is positioned in relation to the catalytic domain to support the intramolecular processing (Guarné *et al.*, 1998).

Using amino acid sequence alignments, it could be shown that picornaviral 3C proteinases belongs to the family of sub-class (trypsin-like) cellular serine proteinases, indicating a chymotrypsin-like fold (Ryan & Flint, 1997 and references therein). The catalytic triad was predicted to be (according to the nomenclature presented here): His^{46}-Asp^{84}-Cys^{163}. The most interesting finding was that of the nucleophilic residue being a cysteine residue, rather than a serine residue (Gorbalenya *et al.*, 1989b). This catalytic triad was confirmed for FMDV with site-directed mutagenesis (Grubman *et al.*, 1995). The crystal structures for human rhino virus-14 (Matthews *et al.*, 1994) and hepatitis A virus (Allaire *et al.*, 1994; Bergmann *et al.*, 1997) 3C proteinases confirmed the chymotripsin-like fold for the picornaviral 3C proteinases.

It is a well established fact that RNA viruses, due to their error-prone RNA replication, have a high mutability rate and therefore have a high level of genomic variation (Holland *et al.*, 1982; Domingo *et al.*, 1990). Genetic characterization of the capsid proteins, especially the VP1 protein, has been the focus of attention. This protein contains important determinants for FMDV neutralization, the G-H loop and C-terminus regions and has been established as a useful tool in the determination of genetic relationships. Previously, it could be shown with amino acid sequence comparisons of the capsid proteins (Palmenberg, 1989) that the South African Territories (SAT) type viruses are genetically different from the European (A, O and C) types. These viruses also show a higher level of intratypic variation than the European types as was shown and discussed in Chapter 2.

Genetic characterization of the non-structural regions of the genome, specifically the Leader and 3C proteinases, are limited and restricted to only a few isolates (Ryan & Flint, 1997). To understand the gene heterogeneity of the Leader and 3C proteinases and their potential impact on the construction of chimeric viruses, representatives of the six FMDV serotypes that occur on the African continent (A, O, C, SAT 1, 2 and 3), were investigated.

3.2 Materials and Methods

3.2.1 Viruses investigated

A list of viruses used in this study is given in Table 3.1. Those indicated in bold were used to generate additional sequencing data needed for the investigation. Viral isolations were performed on oesophageo-pharyngeal specimens (probangs) according to standard procedures at the Onderstepoort Veterinary Institute – Division for Exotic Diseases. These probangs were obtained from buffalo (*Syncerus caffer*) during routine surveys in game parks in southern and eastern Africa. Additional viruses were obtained from the World Reference Laboratory at Pirbright as well as the Botswana Vaccine Institute in Gabarone, Botswana.

Table 3.1: Description of viruses used to assess genetic heterogeneity in the Leader and 3C proteinases

Viruses	Country of origin	Specie	Year of isolation	GenBank Acc. No. (L)	GenBank Acc. No. (3C)
A_{12}	United Kingdom	Bovine	1932	M10975	M10975
A_{10}*	Holland	Bovine	1942	**AF283430**	NA
A_{22}	USSR	Bovine	1965	X74812	X74812
K/37/84/A*#	Kenya	Bovine	1984	**AF283433**	**AF283447**
KEN/1/76/A*	Kenya	Bovine	1971	**AF283432**	NA
GAM/51/98/A*	Gambia	NA	1998	**AF283431**	**AF283446**
O_1Kaufbeuren	Germany	Bovine	1966	X00871	X00871
O_1Campos	Brazil	Bovine	1958	M95781	NA
O_5India*	India	Bovine	1962	**AF283435**	**AF283449**
KEN/1/91/O	Kenya	Bovine	1991	NA	**AF283450**
BKF/2/92/O*	Burkina Faso	Bovine	1992	**AF283434**	**AF283448**
C_1Oberbayen	Germany	NA	1966	X00130	NA
C_1Noville	Switzerland	NA	1965	L19600	**AF283451**
C_1-S8	Spain	Swine	1970	L19601	NA
C_3Resende*	Brazil	Bovine	1965	NA	**AF283452**
ZIM/3/88/1*	Zimbabwe	Buffalo	1988	**AF283437**	**AF283454**
ZIM/3/96/1‡	Zimbabwe	Buffalo	1996	**AF283438**	**AF283455**
KNP/196/91/1‡	South Africa	Buffalo	1991	**AF283436**	**AF283453**
UGA/1/97/1‡	Uganda	Buffalo	1997	**AF283439**	**AF283456**
TAN/1/99/1	Tanzania	Bovine	1999	NA	**AF283457**
KEN/3/57/2	Kenya	Bovine	1957	AJ251473	AJ251473
ZIM/7/83/2†	Zimbabwe	Bovine	1983	**AF283442**	**AF283459**
KNP/19/89/2‡	South Africa	Buffalo	1989	**AF283440**	**AF283458**
ZIM/14/90/2‡	Zimbabwe	Buffalo	1990	**AF283441**	NA
BEC/1/65/3*	Botswana	Bovine	1965	**AF283443**	**AF283460**
KNP/10/90/3‡	South Africa	Buffalo	1990	**AF283444**	**AF283461**
ZAM/4/96/3*	Zambia	Buffalo	1996	**AF283445**	NA

NA – Not available

*World Reference Laboratory, Pirbright, UK

†Botswana Vaccine Institute, Gabarone

‡Onderstepoort Veterinary Institute – Exotic Diseases Division

#Nomenclature used by Embakasi Laboratory, Nairobi, Kenya

3.2.2 RNA extraction, cDNA synthesis and PCR amplification

A modified guanidinium thyocianate/silica method from Boom and co-workers (1990) was used for the rapid extraction of viral RNA. To facilitate the amplification of the region encoding the Leader proteinase, the viral RNA was reverse transcribed using AMV-RT (Promega) and an oligonucleotide (5'-CTACCAGTATTGCCAGATTG-3') which targets the 5' region of VP4 (VP4D). The Leader coding region was subsequently amplified with the VP4D and NCR1 oligonucleotides (Table 2.1), using DynaZyme (Finnzymes). cDNA synthesis for the 3C genomic region was performed using an oligonucleotide situated in the 3D region, binding to nucleotides 6287 to 6304 on KEN/3/57/2 (AJ251473) (3Cminus): 5'-CGCTCTTCMACATCTCT-3'. Amplification of the 3C genomic region was obtained with 3Cminus and an oligonucleotide situated in the 3B region, binding to nucleotides 5574 to 5593 on KEN/3/57/2 (AJ251473) (3Cplus: 5'-CCRGTGAAGAAGCCTGTCGC-3'), again using DynaZyme. Oligonucleotides, 3Cminus and 3Cplus, were designed using nucleotide sequences of A_{10} (X00429), A_{12} (M10975), A_{22} (X74812), O_1 (X00871) and KEN/3/57/2 (AJ251473). Reactions were carried out in the presence of 1.5 mM $MgCl_2$ and nucleotide and oligonucleotide concentrations of 0.2 mM and 0.5 pmol, respectively. After an initial denaturing step of 2 minutes at 94°C, thirty cycles of denaturing at 95°C for 45 sec, followed by an annealing step at 55°C for 30 sec and an elongation step at 72°C for 90 sec, were performed. Altered annealing temperatures for a few isolates had to be used to refine amplification. The resulting fragments were analysed by agarose gel electrophoresis and recovered from the agarose gel with Cleanmix kit (Talent) according to the specification of the manufacturer to be characterized molecularly.

3.2.3 Nucleotide sequence determination and analysis

The nucleotide sequence of the 750bp (Leader proteinase) and 680bp (3C proteinase) fragments were determined using the ABI PRISM 377 DNA Sequencer from Perkin Elmer Applied Biosystems and T7 DNA polymerase (sequenase version 2.0, Amersham). Analysis of sequencing data was performed with the DAPSA version 2.9 program (Harley, 1994). Amino acid sequence alignment of the 3C proteinases with hepatitis A 3C proteinase (GenBank Acc. No.X70845) was done with CLUSTALW.

All nucleotide sequences determined in this study have been submitted to GenBank under the accession numbers indicated in Table 3.1.

3.2.4 Determination of gene relationships for FMDV Lab and 3C proteinase coding regions

The analysis for the Lab^{pro} and $3C^{pro}$ coding regions was carried out on the full-length nucleotide region (L=618 nt; 3C=639 nt). The corresponding regions of equine rhinitis type A virus (GenBank Acc. No. X96870) and hepatitis A virus (GenBank Acc. No. X70845) were selected as outgroups for Lab^{pro} and $3C^{pro}$ analysis, respectively. Gene relationships were established for the full data set (Lab^{pro}: N=25 and $3C^{pro}$: N=21) using the neighbor-joining method (*p*-distances) included in MEGA version 1.0 (Kumar *et al.*, 1993).

3.2.5 Structural modeling of FMDV Lb proteinase

ZIM/7/83/2 Lb amino acid sequence was aligned with that of O_1Kaufbeuren with ClustalX. This alignment was used for homology modeling of the ZIM/7/83/2 structure employing MODELLER version 4.0 (Sali & Blundell, 1993) on a Silicon Graphics Power Indigo2 Extreme. The approach is based on the satisfaction of spatial restraints. The co-ordinates for the crystal structure of O_1Kaufbeuren Lb^{pro} was generously provided by Dr. Alba Guarné (NIH, Washington) and used as template for homology modeling. The modeled structure was verified with PROCHECK version 3.4.4 (Laskowski *et al.*, 1993) and compared to the template with GRASP version 1.3.6 (A. Nicholls, Columbia University) and WHAT-IF (Vriend, 1990).

3.2.6 Determination of FMDV P1 cleavage sites mediated by the 3C proteinase

Representatives of the six serotypes were chosen to compare the 3C proteinase cleavage sites within the P1 region. Amino acid sequence data were obtained from GenBank with the exception of KNP/196/91/1, determined as discussed in Chapter 2, and KNP/10/90/3 (L.E. Heath, unpublished results).

3.3 Results and Discussion

3.3.1 Gene heterogeneity and gene relationships of Lab proteinases

Twenty four representatives of the six FMDV serotypes that occur on the African continent, were chosen for this study (Table 3.1). Representatives for types A, O and C were taken from Europe, Asia, Brazil, West and East Africa while representatives of the SAT types originate from southern and eastern Africa. As expected, the catalytic triad of the proteinase (Cys52-His149-Asp164), as well as the Lab and Lb forms of the protein (Figure 3.2) are conserved. A difference in length in the Lab form was detected. A type A isolate from Kenya (K/37/84) is 202 amino acids in length, whilst another type A isolate from Kenya (KEN/1/76) is the same length (201 residues) as the European type A, O and C's. All the SAT types are three residues shorter (199), the same length as the type A and O isolates from West Africa.

In Figure 3.2 the secondary structural motifs of Lb^{pro} are given as previously indicated by Guarné and co-workers (1998). Alpha helices 1 and 4 are well conserved, while helices 2 and 3 display a higher level of variation. With the exception of β4, the β-sheets are generally conserved. A high level of variation is observed in the C-terminus extension (CTE), though. The SAT types vary considerably from types A, O and C in this region, although these differences seem to be clustered. The two isolates from East Africa (UGA/1/97/1 and KEN/3/57/2) differ in the CTE region not only from types A, O and C, but also from the SAT type viruses. The cleavage site with VP4 (Lys/Gly) of these two isolates corresponds however with types A, O and C, while that of the southern African SAT types differ: Arg/Gly.

A neighbor-joining analysis depicting Lab proteinase gene relationships (Figure 3.3) between the six different serotypes, generated a phylogenetic tree with 2 major lineages. The first major group (lineage I, blue) comprises of types A, O and C as well as two SAT isolates from East Africa. The European types O and C group according to serotype, while the two isolates from Asia (A_{22} and O_5India) group with a type A isolate (KEN/1/76) from Kenya and a type O (BKF/2/92) isolate from West Africa to form a sub-cluster within lineage I. Another very interesting and unexpected grouping

```
                                            ββ1ββ    ββ2ββ         αααα1ααα
                Lab                         Lb                     *      60
O1Kaufbeuren    MNTTDCFIALVQAIREIKALFLSRTTG-KMELTLYNGEKKTFYSRPNNHDNCWLNAILQL
O1Campos        ......................P.....................................
O5India         .T........LH.L....T......Q.....F.......................T....
BKF/2/92/O      ..................L.LKGLK--....F................N......T....
A12             ....N......H.....R.F....A......F......R................T....
A10Holland      ....N......YL.....T..R.........F..H....................T....
A22             ..........LY.L....FL.....Q.............................T....
K/37/84/A       ...........R....F.I..SL.PLAR...F.....I.........K............
KEN/1/76/A      ..........LR.F....T.....VR.....F............................
GAM/51/98/A     .................RL.LKGSK--....F..H............R.......T....
C1Noville       .........V.N.....R....P........F..HD....V..............T....
C-s8c1          .........V.N..K.VR....P..A.....F..HD....V..............T....
C1Oberbayen     .......T.V.N.....R....P........F..HD....V.......Q......T....
ZIM/3/88/1      .K.....DV.LEIFHRFRQT.K---.DR...F...............T.G.....SL...
ZIM/3/96/1      .K..G..DV.IEIAHRLRQ.NK---.DR...F...............T.G.....SL...
KNP/196/91/1    .K.....SV.FEIFHRLRHT.K---.ER...F......R........K.G.....SL...
UGA/1/97/1      ..................S..R.S---R...F.......................T....
KEN/3/57/2      ..................L..KGI---R..K................S.......T....
ZIM/7/83/2      .K.....NV.LEI.YRFRHT.K---.DR...F...............K.G.....SL...
ZIM/14/90/2     .K.....NV.AEIFHRL.QINK---.DR...F..H............T.G.....SL...
KNP/19/89/2     .K.....DV.FEIFHRFRHT.K---.ER...F...............K.G.....SL...
BEC/1/65/3      .K.....NV.FEIFHRFGQT.K---ADR...F...............T.G.....SL...
KNP/10/90/3     .K.....NV.LETFHRFRNV.K---.DR...F.....D.........T.G.....SL...
ZAM/4/96/3      .K.....D..LE.FHRLRQT.K---.DR...F...............R.G.....SL...
```

```
                αααα    αα2αα       αααααα3αααααα      ααα4ααα         ββ3ββ
                                                                         120
O1Kaufbeuren    FRYVEEPFFDWVYSSPENLTLEAIKQLEDLTGLELHEGGPPALVIWNIKHLLHTGIGTAS
O1Campos        ..............T.............................................
O5India         ....D........D..........R...EI..............................
BKF/2/92/O      ....D........N..........D...KV..............................
A12             ....D........N.......A......E.......................Q.......
A10Holland      ....D........N.......D....VGK.H.....................Q.......
A22             ....D........D......C...R...EI..............................
K/37/84/A       ....D........E......IQ..G...E....D.R........V...............
KEN/1/76/A      ....D....E...D......V...R...E..........................V....
GAM/51/98/A     ....D........D.......Q..E...E..........................V.S..
C1Noville       ....D........N..............E......R........................
C-s8c1          ....D........N..............E......R........................
C1Oberbayen     ....D........N..............E......R........................
ZIM/3/88/1      ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..DC.Q..V..ST
ZIM/3/96/1      ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..DC.Q..V..ST
KNP/196/91/1    ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..GC.Q..V..ST
UGA/1/97/1      ....D........N.......Q..E...E.......................Y.......
KEN/3/57/2      ....D........N.......R..E...E......R................Y.......
ZIM/7/83/2      ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..DC.Q..V..ST
ZIM/14/90/2     ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..DC.Q..V..ST
KNP/19/89/2     ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..DC.Q..V..ST
BEC/1/65/3      ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..DC.Q..V..ST
KNP/10/90/3     ....D..L.ESE.L....K..DM..R.S.Y.K.D.SD.......L.L..DC.Q..V..ST
ZAM/4/96/3      ....D..L.ESE.L....K..DM....S.Y.K.D.SD.......L.L..DC.Q.....ST
```

```
                    ββ4ββ        βββ5βββ    βββ6βββ/ββ7βββ  βββ8ββ      βββ9
                                            *                *              180
O1Kaufbeuren        RPSEVCMVDGTDMCLADFHAGIFLKGQEHAVFACVTSNGWYAIDDEDFYPWTPDPSDVLV
O1Campos            ..............................................................
O5India             ..........................................K...................
BKF/2/92/O          ..........................H...................................
A12                 ..........................E...................................
A10Holland          ..........S............M.............D.......................
A22                 ..................................D.......................
K/37/84/A           ..................................E.......................
KEN/1/76/A          ......................V...............................
GAM/51/98/A         ..........................V.......L...........E.............
C1Noville           ........................M.....................................
C-s8c1              ........................M..R..................................
C1Oberbayen         ........................M.....................................
ZIM/3/88/1          ....I.VIN.VT.T.........I..T......LN..E........V........EN..A
ZIM/3/96/1          ....I.VIN.VV.T.........I..T......LN..E........V........EN..A
KNP/196/91/1        ....I.VIN.VT.T.........I..T......LN..E........V........EN..A
UGA/1/97/1          ..........................T.......L..D.....................C
KEN/3/57/2          ........................M..T......LL..D.....................C
ZIM/7/83/2          ....I.VIN.VV.T.........I..T......LN..E........V........EN..A
ZIM/14/90/2         ....I.VIN.VT.T.........I..T......LN..E........V........EK..A
KNP/19/89/2         ....I.VIN.VT.T.........I..T......LN..E........V........EN..A
BEC/1/65/3          ....I.VIN.VV.T.........I..T......LN..E........V........EN..A
KNP/10/90/3         ....I.VIN.VV.T.........I..T......LN..E........V........EN..A
ZAM/4/96/3          ....I.VIN.VV.T.........I..T......LN..E........V........EN..A

                    βββ↛CTE
                                                                VP4
O1Kaufbeuren        FVPYDQEPLNGEWKAKVQRKLK 201                  GAGQ
O1Campos            ...................... 201                  ....
O5India             ..................KR.. 201                  .T.R
BKF/2/92/O          ..............T...KR.. 199                  ....
A12                 ............G...N...... 201                  ....
A10Holland          ............D..TL...... 201                  ....
A22                 ..................KR.. 201                  ....
K/37/84/A           ..............T...K.... 202                  ....
KEN/1/76/A          ...................R.. 201                  ...H
GAM/51/98/A         .........S....T...K.... 199                  ...R
C1Noville           ..........EG...N...... 201                  ....
C-s8c1              ..........EG...S...... 201                  ....
C1Oberbayen         ..........EG...N...... 201                  ....
ZIM/3/88/1          Y........DVD.QDRAGLF.R 199                  ....
ZIM/3/96/1          Y........DVD.QDRAGLF.R 199                  ....
KNP/196/91/1        Y........DVD.QGRAGLF.R 199                  ....
UGA/1/97/1          .....M..F..DAIV.ATAYV. 199                  ....
KEN/3/57/2          .....M..F..SSI..ATAYV. 199                  ....
ZIM/7/83/2          Y........DVD.QDRAGLF.R 199                  ...H
ZIM/14/90/2         Y........DVD.QDRAGLF.R 199                  ....
KNP/19/89/2         Y........DVD.QDRAGLF.R 199                  ....
BEC/1/65/3          Y........DVD.QDRAGLF.R 199                  ....
KNP/10/90/3         Y........DVD.QDRAGLF.R 199                  ....
ZAM/4/96/3          Y........DVD.QDRAGLF.R 199                  ....
```

Figure 3.2: Amino acid sequence alignment of Leader proteinases. Both the Lab and Lb forms of the proteinase are indicated as well as the first four amino acid residues of VP4. The catalytic residues are indicated with an asteriks (*). Secondary structural motifs (α-helices and β-sheets are indicated according to Guarné and co-workers (1998)).

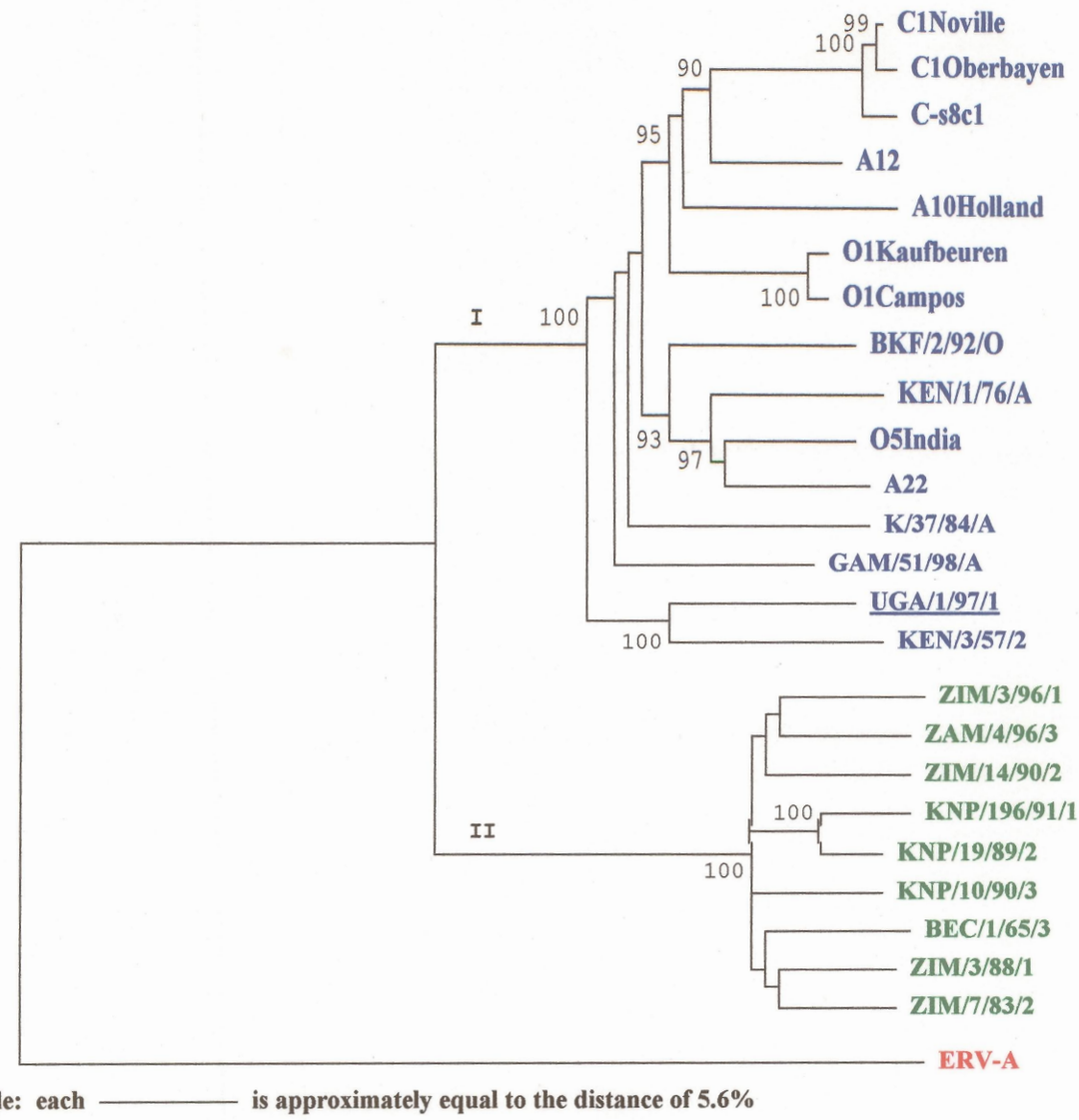


Figure 3.3: Neighbor-joining tree (*p*-distances) depicting Lab^{pro} gene relationships. Bootstrap values > 90 and based on 1000 bootstrap replications are shown with the equine rhinitis virus type A (ERV-A) Leader proteinase selected as outgroup. The two different lineages (I and II) are indicated in blue and green, respectively.

in this lineage, is that of the two East African SAT types (UGA/1/97/1 and KEN/3/57/2). Lineage II (green) contains all the southern African SAT types with no apparent phylogenetic sub-structuring, except for the grouping of a SAT 1 isolate (KNP/196/91) and a SAT 2 isolate (KNP/19/98) from the Kruger National Park, South Africa. The amino acid variation in lineage I (36%) was found to be much higher than the variation in lineage II (10%). This observation is unexpected, as the structural proteins of the SAT types are two to three times more variable than that of types A, O and C. Overall amino acid identity for the 24 viruses investigated, amounts to 58%.

3.3.2 Structural model for ZIM/7/83/2 Lb proteinase

The impact of the amino acid differences between lineages I and II (Figure 3.3) on the structure of the Leader proteinase, was investigated by constructing a three-dimensional model for a representative of lineage II (ZIM/7/83/2) (F. Joubert, Dept. Biochemistry, UP). A comparison with the previously elucidated crystal structure of O_1Kaufbeuren Leader proteinase is given in Figure 3.4. Despite an identity on amino acid level of only 64% between the two proteinases, the three-dimensional fold is completely conserved. The only difference between the two structures appears to be in the C-terminus extension (CTE). It is evident from amino acid comparisons in this region (Figure 3.2) that a high level of substitutions is present between the two lineages. This was also observed for the two East African SAT isolates (UGA/1/97/1 and KEN/3/57/2) within lineage I.

Previously RKLK↓GAGN was established to be the cleavage site of L^{pro} from types A, O and C. Results generated in this study indicated that the corresponding cleavage site for L^{pro} from southern African SAT types differs from the other types and is cleaved at: LFLR↓GAGN. The processing site for UGA/1/97/1 was found to correspond with the previously determined KEN/3/57/2: AYVK↓GAGN. In general these differences in the CTE region are not expected to have major implications on the three-dimensional structure of the proteinase, which are confirmed with the ZIM/7/83/2 L^{pro} model. A possible exception is the presence of the phenylalanine residue in the SAT isolates from

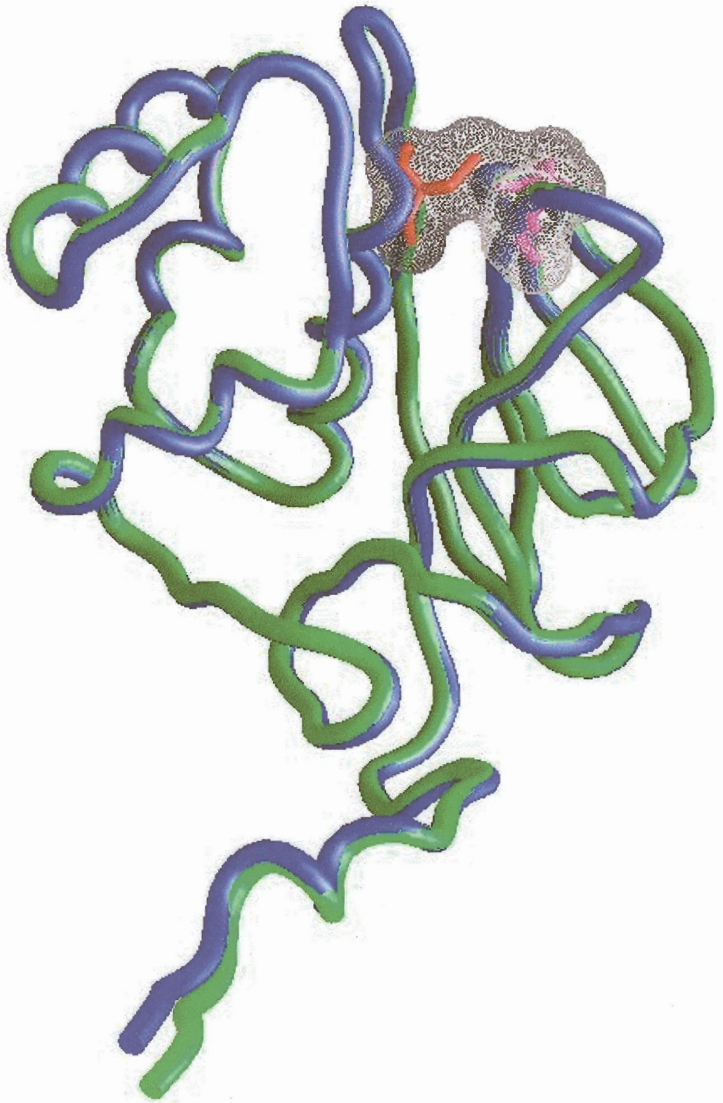


Figure 3.4: Superimposed C_α traces of O_1Kaufbeuren Lb^{pro} (blue) and ZIM/7/83/2 Lb^{pro} (green). The catalytic Cys-51 (red) and His-148 (purple) residues, which are situated on top of the deep cleft, are indicated as well as the electron density displayed by these residues.

southern Africa (T. Skern and A. Guarné, personal communication). The effect (if any) of these differences on self-processing are now investigated using biological assays.

3.3.3 Gene heterogeneity and gene relationships of 3C proteinase

Representatives (21 isolates) of serotypes A, O, C and SAT 1, 2 and 3 were chosen to investigate the gene heterogeneity in the 3C proteinase (Table 3.1). Amino acid sequence alignment of the isolates shows the complete conservation of the active site residues (His46-Asp84-Cys163), as expected (Figure 3.5). Alignment of the FMDV $3C^{pro}$ regions with that of hepatitis A virus (HAV) using CLUSTALW also shows the conservation of the active site residues. The three-dimensional structure of HAV (Allaire *et al.*, 1994; Bergmann *et al.*, 1997) indicates a chymotripsin-like fold, comprising of two anti-parallel β-barrels – each barrel consisting of six β-sheets. The secondary structural motifs as previously predicted for picornavirus 3C proteinases (Gorbalenya *et al.*, 1989) are indicated for the FMDV isolates in Figure 3.5 and shows the conservation of the 12 β-sheets. Overall the sheets are well conserved, although some differences can be seen in βE, βG and βJ. The C- and N-terminus α-helices are well conserved.

A neighbor-joining tree depicting $3C^{pro}$ gene relationships between the different serotypes is shown in Figure 3.6. Similar phylogenetic groupings were obtained for $3C^{pro}$ as was observed for Lab^{pro}. The group indicated in blue is inclusive of isolates form types A, O and C from Europe, Asia, West and East Africa as well as a single isolate from Brazil. In this analysis, it was interesting to observe that the SAT 2 virus KEN/3/57 grouped with two isolates from West Africa (GAM/51/98/A and BKF/2/92/O) and forms a sub-cluster within the first lineage. The group in green (Figure 3.6) represents SAT viruses from southern Africa. UGA/1/97/1 grouped with lineage II during this analysis instead of lineage I as previously found (Figure 3.3). This was the only difference obtained for the Lab^{pro} and $3C^{pro}$ phylogenetic analyses. Again, amino acid variation for lineage I (16%) was found to be higher than for lineage II (12%).

```
                    αaCαα   ββAββ          βββBββ ββββCββββ     βDββ
                                                          *          60
O1Kaufbeuren SGAPPTDLQKMVMGNTKPVELILDGKTVAICCATGVFGTAYLVPRHLFAEKYDKIMVDGR
O5India      ....................................S.................L...
BKF/2/92/O   ............A...........I.............................I...
KEN/1/91/O   ....................................................L...
A12          ....................................................L...
A22          ....................................................L...
K/37/84/A    ....................................................L...
GAM/51/98/A  ............A.............V.........................I...
C1Noville    ....................................V....G..........L...
C3Resende    ....................................................L...
KNP/196/91/1 ..C.........A.V...........L.........................L...
ZIM/3/88/1   ..C.........A.V...........L.........................L...
ZIM/3/96/1   ..C.........A.V...........L.........................L...
TAN/1/99/1   ..C.........A.V...........L.........................L...
UGA/1/97/1   ............A.V.........I..L.........................L...
KEN/3/57/2   ............A.............................................I...
ZIM/7/83/2   ............A.V...........L.........................L...
KNP/19/89/2  ..C.........A.V...........L.........................L...
KNP/10/90/3  ..C.........A.V...........L.........................L...
BEC/1/65/3   ..C.........A.V...........L.........................L...
HAV          .TLEIAG.VRKNLVQFGVE.--KN.CVRWVMN.L..KDDWL...S.AYKFEK.YE.MEFY
             αααCααα  βββA1ββ        βββB1βββββ  ββC1ββ          βββD1

                    βββββEββββ          βββFβββ                  βββGββ
                                           *                        120
O1Kaufbeuren AMT-DSDYRVFEFEIKVKGQD-MLSDAALMVLHRGNRVRDITKHFRDTARMKKGTPVVGV
O5India      ....................................................................
BKF/2/92/O   .....R.F..................V...........A.....C.Q...R.........
KEN/1/91/O   ....................................................................
A12          ....................................................................
A22          ....................................................................
K/37/84/A    ....................................................................
GAM/51/98/A  .....R.F.....................................Q.K.R.........
C1Noville    .L.....................................................V.............
C3Resende    ..................................T...........V.............
KNP/196/91/1 .L.....F......V..................S......L.G.....MKLS..S.....
ZIM/3/88/1   .LK....F......V..................S......L.G.....MKLS..S.I...
ZIM/3/96/1   .L.....F......V..................S......L.G.....MKLS..S.I...
TAN/1/99/1   .L.....F......V..................T......L.G.....MKLS..S.....
UGA/1/97/1   .L..NG.F......V.................N..Q......A.....V.VA..N.....
KEN/3/57/2   .....R.F.....................................R....Q...R.........
ZIM/7/83/2   .L.....F......V..................S......L.G.....MKLS..S.....
KNP/19/89/2  .L.....F......V..................S......L.G.....MKLS..S.....
KNP/10/90/3  .L...G.F......V..................S......L.G.....MKLS..S.....
BEC/1/65/3   .L.....F......V..................S......L.G.....MKLS..S.....
HAV          FNRGGTYYSISAGNVVIQSL.VGFQ.VV..KVPTIPKF..I.Q..IKKGDVPRALNRLAT
             ββ   βE1aβ     βE1bβ      βF1β                        ββA2
```

```
                β          βββββH     ββ                    ββIβ       βββββJββββ
                                                                        *   180
O1Kaufbeuren INNADVG--RLIFSG-----EALTYKDIVVCMDGDTMPGLFAYRAATK--AGYCGGAVLA
O5India      ..........................................P.................
BKF/2/92/O   ..........................................................
KEN/1/91/O   ..........................................................
A12          V.....................................S....K.................
A22          .....................................K.....................
K/37/84/A    ..........................................................
GAM/51/98/A  .........K.................................................
C1Noville    .....................................K.....................
C3Resende    .....................................K.....................
KNP/196/91/1 V.................D......L.................G....V....A....
ZIM/3/88/1   V.................D......L.................G....V....A....
ZIM/3/96/1   V.................D......L.................G....V....A....
TAN/1/99/1   V.................D......L.................G....V....A....
UGA/1/97/1   V.................D.......................G....V....A...T
KEN/3/57/2   V....RW..................................K.....................
ZIM/7/83/2   V.................D......L.................G....V....A....
KNP/19/89/2  V.................D......L.................G....V....A....
KNP/10/90/3  V.................D......L.................G....V....A....
BEC/1/65/3   V.................D......L.................G....V....A....
HAV          LVTTVNGTPM..SE.PLKMEERA..VHKKNDGTTVDLTVDQ.WRGKGEGLP.M..G.LVS
             ββββ  βββββB2βββββββ ββaβ     ββbβ  ββC2ββ          βD2β
```

```
                β       βββββKββββ/βLβ  ααaNααα
O1Kaufbeuren KDG-ADTFIVGTHSAGGNGVGYCSCVSRSMLLKMKAHIDPEPHHE   213
O5India      ..........................................L..   213
BKF/2/92/O   .....E.....................K....Q....V.......   213
KEN/1/91/O   .....................................V.......   213
A12          ..........................K....R....V....Q..   213
A22          ..........................................    213
K/37/84/A    .....................................MH......   213
GAM/51/98/A  ....PE.........................Q....V..D..S.   213
C1Noville    ..........................................    213
C3Resende    ....PE......................................   213
KNP/196/91/1 .....K.V.......................Q.......P..T.   213
ZIM/3/88/1   .....K.V.......................Q.......P..T.   213
ZIM/3/96/1   .....K.V.......................Q.......P..T.   213
TAN/1/99/1   .....K.V.......................Q.......P..T.   213
UGA/1/97/1   .S..SQ.V.I..................K...DQ.......A..T.   213
KEN/3/57/2   .....E.....................K....Q....V.......   213
ZIM/7/83/2   .....K.V.......................Q.......P..T.   213
KNP/19/89/2  .....K.V.......................Q.......P..T.   213
KNP/10/90/3  .....K.V.......................Q.......P..T.   213
BEC/1/65/3   .....K.V.......................Q.......P..T.   213
HAV          SNQSIQNA.L.I.V....SILVAKL.TQE.FQNIDKKIESQ----   219
                    ββE2ββ       βF2β  αNα
```

Figure 3.5: Amino acid sequence alignment of FMDV $3C^{pro}$ indicating differences between southern African types and types A, O and C. The conservation of the putative catalytic residues (His-46; Asp-84; Cys-163) in the isolates investigated as well as hepatitis A virus (HAV) , is indicated with an asteriks (*). Secondary structural motifs for FMDV as previously predicted by Gorbalenya and co-workers (1989) are indicated as well as for HAV (Ryan & Flint, 1997).

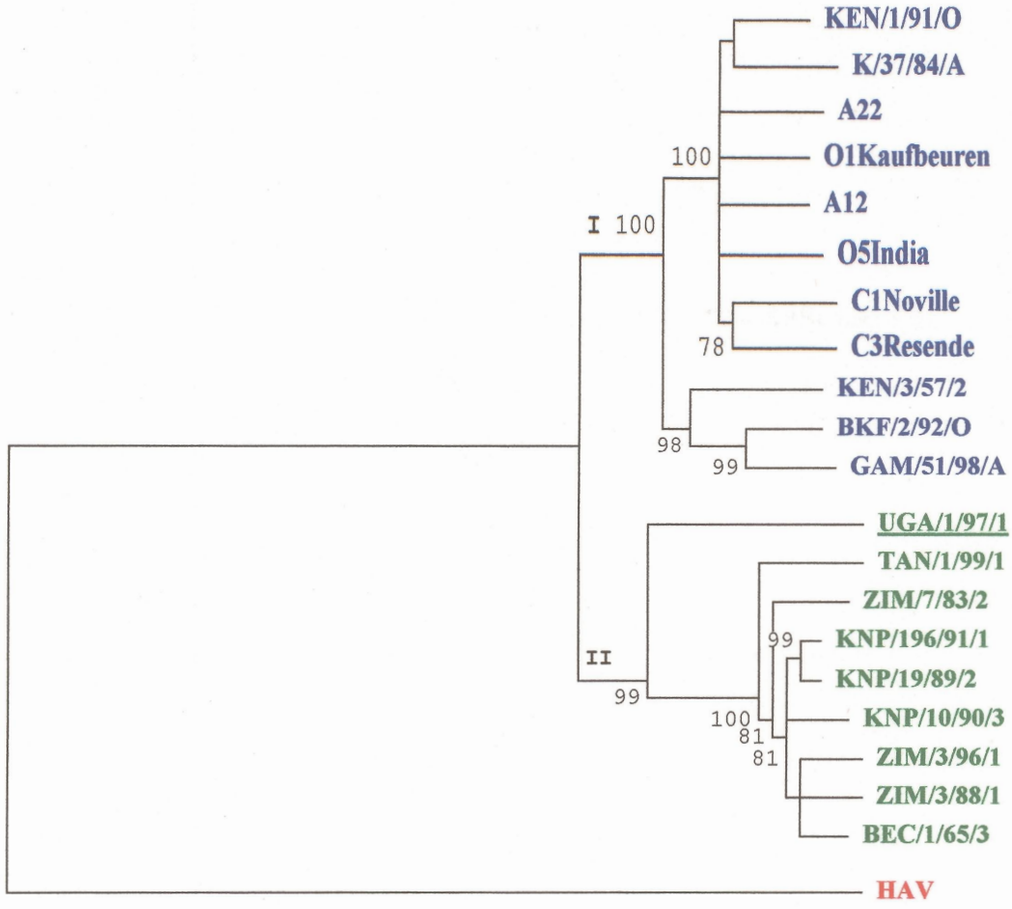


Figure 3.6: Neighbor-joining tree (*p*-distances) depicting $3C^{pro}$ gene relationships. Bootstrap values > 75 and based on 1000 replications are indicated with hepatitis A (HAV) 3C proteinase selected as outgroup. Lineage I is presented in blue and lineage II in green.

3.3.4 Comparison of FMDV P1 cleavage sites mediated by the 3C proteinase

The polyprotein precursor is proteolytically processed by viral proteinases. The primary cleavage site of the FMDV $3C^{pro}$ is at the 2B/2C junction (Ryan & Flint, 1997). The proteinase is, however, also responsible for a series of secondary cleavages. Preferred processing sites of the $3C^{pro}$ for another picornavirus member, poliovirus, are Gln-Gly (Q/G) (Bazan & Fletterick, 1988). Inspection of the P1 cleavage sites of the different FMDV serotypes (excluding Asia-1), revealed variations between the different types and poliovirus (Figure 3.7). The Q/G processing site is conserved in the SAT types for the VP0/VP3 junction, but is replaced by an acidic glutamic acid (E) residue in types A, O and C. The equivalent of this site in HAV and human rhino virus type 14 (HRV14) is Q/M and Q/G, respectively (Seipelt *et al.*, 1999). The Gln residue is conserved in the VP3/VP1 site in all the serotypes except type O, where it is again replaced by a glutamic acid. The Gly residue is however replaced in the FMDV types by a more bulky Thr residue. The corresponding VP3/VP1 site in HAV is Q/V and in HRV14 E/G (Seipelt *et al.*, 1999). In the VP1/2A junction site, the Gln residue in completely conserved. Variation occurs however on the 2A side of the junction site. Type O contains the neutral and polar Thr residue, but this residue is replaced in the other types by a hydrophobic Leu (L) residue. BEC/1/65/3 is an exception and contains a methionine residue (also hydrophobic) instead of Thr in this position.

The recognition of these processing sites is dependent on their position within the polyprotein. The regions flanking these sites might therefore play an important role in the recognition process. Although it is evident from Figure 3.7 that these regions are not highly conserved between the European and SAT types, one would assume the three-dimensional structure of these regions to be conserved. The efficient conservation of the proline residue in the –4 position from the cleavage site, may be of importance. The structural role this residue might play, due to its aromatic nature has been implicated before for both FMDV and encephalomyocarditis (EMCV) (Palmenberg *et al.*, 1984). Exceptions observed in this study entail the presence of an alanine residue in this position for types O and C at the VP3/VP1 cleavage site as well as the occurrence of a valine residue in the –4 position at the VP1/2A cleavage site for

	VP0 VP3	VP3 VP1	VP1 2A	
O1Kaufbeuren	NVHVAGEFPSKE GIFPVACSDG	ELRLPVDARAE TTSAGESADPV	KQKIVAPVKQ TLNFDLLKLAG	X00871
O1Campos	..........E G.....	.....E T.....	.....Q T.....	M95781
A12	Y.....E G.....	.....I.P.SQ T.AT.....	....I..G.Q L.....	M10975
A22	H.....L...E G.V.....	.....I.P.SQ T..T.....	....I..A.Q L.....	X74812
C1Oberbayen	..Y....L...E G..S.....	.....QQ T.TT.....	..PL...A.Q L.....	X00130
C3Resende	.....L...E G.....-...	.....LQ T.TT.....	..RLI..A.Q LS.....	M90381
BOT/1/68/1	..Y....K.V.Q G.L...V...	.F.M.ISPSRQ T.....G....	.TTL.K.A.Q LS.....	Z98203
KNP/196/91/1	..Y....K.A.Q G.L...V.V.	.F.M.ISPSRQ T.....G.E..	RTA.TK...Q LC.....	AF283429
KEN/3/57/2	..Y....L.G.Q G.V....A..	.F.F.I.PVRQ T.....G.EV.	FDAPIGVA.Q L.....	AJ251473
RHO/1/48/2	..F....M.A.Q G.I....A..	.F.F.I.PVRQ T.AV..G..V.	FDAPIGVE.Q LF.....	AJ251475
ZIM/7/83/2	..F....K.A.Q G.....F..	.F.F...PVRQ T..S..G..V.	FDSPIGVE.Q LC.....	AF136607
BEC/1/65/3	..YR...R...Q G.....N..	.....INPATQ T.....G..V.	.TPL.K.D.Q MC.....	M28719
KNP/10/90/3	..Y....K.T.Q G.V....H..	.F...INPV.Q T.....GG.V.	.T.L...D.Q LC.....	**Unpublished results**

Figure 3.7: P1 cleavage sites (VP0/VP3; VP3/VP1; VP1/2A) processed by $3C^{pro}$ are indicated as well as regions adjacent to the cleavage site. Amino acids identical to the cleavage site of poliovirus are indicated in blue, while those that differ are indicated in red. GenBank Accession numbers (except KNP/10/90/3) are given.

SAT 2 (Figure 3.7). However, in both cases a proline residue is present in the –7 position.

3.4 Conclusions

The genetic heterogeneity of the L^{pro} and $3C^{pro}$ regions of representatives of six FMDV serotypes originating from different geographical localities was investigated. The study revealed that the L^{pro} and $3C^{pro}$ genomic regions of the SAT type viruses from southern Africa are distinct from that of types A, O and C. Interestingly, results indicated that the amino acid sequences of L^{pro} and $3C^{pro}$ of the SAT types are less variable than the corresponding amino acid sequences of types A, O and C. This is unexpected as it has been shown previously that the structural proteins of SAT 2 are more variable than that of types A, O and C (see Chapter 2). Phylogenetic analysis of the L^{pro} and $3C^{pro}$ genomic regions revealed similar structuring, although different from what was previously obtained for the structural proteins (Bastos, 1998). In the latter case, isolates grouped strictly according to serotype.

An interesting observation made during analysis of the L^{pro} and $3C^{pro}$ gene relationships is the different groupings obtained for the East African strain, UGA/1/97/1. This strain was found to group with lineage I with L^{pro} analysis, but with lineage II with $3C^{pro}$ analysis. These findings strongly suggest the occurrence of viral recombination. This phenomenon has previously been shown in different laboratories for both poliovirus (Ledinko, 1963; King, 1988) and FMDV (Pringle, 1965; King *et al.*, 1985; Giraudo *et al.*, 1988). Due to the high number of recombination sites detected in the RNA genome of FMDV, it was concluded that recombination is a general process (King *et al.*, 1985). Despite these findings, no reports, as far as could be determined, on the occurrence of FMDV recombination in nature had previously been reported. The SAT 1 Ugandan strain was isolated from buffalo (Table 3.1) in the Queen Elizabeth National Park in Uganda. As multiple viral infections for individual buffaloes occur frequently in nature (Hedger, 1972), it could create the ideal physical conditions for viral recombination to occur.

The heterogeneity in the L^{pro} and $3C^{pro}$ genomic regions emphasize, yet again, the distinctiveness of the SAT type viruses. Nevertheless, it was shown in this study that the overall three-dimensional fold for L^{pro} is completely conserved. The biological significance of the observed differences in the CTE region will subsequently be determined in a biological assay system. Together with the heterogeneity observed in the $3C^{pro}$ P1 cleavage sites as well as the regions adjacent to the cleavage sites, these differences could possibly impact on the construction of viable chimeric viruses between different types, e.g. types A and SAT 2. This aspect will be addressed in the following chapter.