CHAPTER 2

Molecular characterization of the structural-protein-coding region of ZIM/7/83/2, a SAT 2 type foot-and-mouth disease virus[1]

2.1 Introduction

Six of the seven serotypes of FMDV (South African Territories (SAT) 1, 2 and 3 as well as serotypes A, O and C), occur in sub-Saharan Africa. Sporadic reports of the occurrence of the SAT types outside this region have been documented before. These include the detection of SAT 1 in Northern Africa and the Middle East (Pereira, 1981) as well as the recent outbreak of SAT 2 in Saudi Arabia. Despite this, the SAT types occur almost exclusively in sub-Saharan Africa. Types A, O and C occur frequently in Western, Eastern and Central Africa where they are endemic in cattle populations and unlike the SAT types, are not established in wildlife (Hedger *et al.*, 1973). Historically, SAT 2 serotype has been responsible for more FMD outbreaks (48%) in domestic animals in southern Africa than any other type (Thomson, 1994). The SAT 2 serotype was therefore selected to be employed together with the type A_{12} genetic backbone in the construction of a recombinant chimeric virus. Towards this objective, the structural-protein-coding (P1) region of the SAT 2 vaccine strain, ZIM/7/83, was amplified using RT-PCR from partially purified virus, cloned into an appropriate vector, molecularly characterized and compared with known FMDV SAT 2 isolates.

2.2 Materials and Methods

2.2.1 Viral and bacterial strains

The FMD vaccine strain, ZIM/7/83/2 (passage history: B1BHK5B1), was used during this study. This bovine outbreak strain originates from western Zimbabwe and was obtained from the Botswana Vaccine Institute in Gabarone. RHO/1/48/2 (passage

[1] Parts of the results presented here have been published in *Virus Genes* 1999; 19 (3): 229 – 233.

history: BTY2RS2), also a bovine outbreak strain, originates from Zambia and was obtained from Dr. Nigel Ferris at the World Reference Laboratory for Foot-and-mouth Disease, Pirbright, UK. *Escherichia coli* JM109 (genotype: *rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17 (r_K-,m_K+), *rel*A1, *sup*E44, delta(*lac-pro*AB), [F', *tra*D36, *pro*AB, *lac*I^qZdeltaM15]) was used for transformation experiments and was obtained from Promega.

2.2.2 *Oligonucleotides*

All oligonucleotides used in this study are summarized in Table 2.1. The specific region on the FMDV genome where the oligonucleotides bind, are indicated according to the KEN/3/57/2 (AJ251473) sequence, unless indicated differently. NCR1 was designed from the full-length nucleotide sequences of A_{10} (X00429), A_{12} (M10975), A_{22} (X74812), O_1 (X00871) and KEN/3/57/2 (AJ251473).

2.2.3 *Partial purification of viruses and RNA extraction*

Viruses were twice passaged on "Instituto Biologico Rim Suino" or IB-RS-2 cells (a pig kidney cell line), first at a low multiplicity of infection (m.o.i), followed by a high m.o.i. A 30% (w/v) sucrose gradient together with high-speed ultracentrifugation (150 000 x *g* for 3h at 4°C) was used to pellet the viral particles. RNA was extracted from the pellet which was resuspended in 1 x TE and sonicated for 3 to 5 min. Sodium dodecyl sulfate (SDS) and 2-mercaptoethanol (1% each) were added to the sonicated pellet followed by a phenol:chloroform (1:1) extraction. This was followed by phenol-(1:1) and diethylether (2:1) extractions. The supernatant was subsequently precipitated with 1/10 volume 3M sodium acetate pH 5.2 and 2 volumes ice cold 100% ethanol for 2 h at -20°C, dried *in vacuo* and stored at -70°C (Vosloo, 1992).

2.2.4 *cDNA synthesis*

Using AMV-RT (Promega), the viral RNA was reverse transcribed with the 2B208R oligonucleotide. This primer targets the 2B region of the genome (Table 2.1; Figure 2.1). The reaction was carried out at 42°C for 2h in the presence 2% DMSO.

Table 2.1 Summary of oligonucleotides

Oligonucleotide	Sequence	Purpose	Reference	Binding nucleotides
2B208R (20 mer)	5'-ACAGCGGCCATGCACGACAG-3'	cDNA synthesis	N. Knowles (pers. comm.)	3753 – 3772
NCR1 (22 mer)	5'-TACCAAGCGACACTCGGGATCT-3'	PCR amplification and nucleotide sequencing	This study	549 – 570
WDA (21 mer)	5'-GAAGGGCCCAGGGTTGGACTC-3'	PCR amplification and nucleotide sequencing	Beck & Strohmaier (1987)	3531 – 3551
WUS (23 mer)	5'-CCACRTATTACTTYTGTGACCTG-3'	Nucleotide sequencing	A.D.S. Bastos (pers. comm.)	3055 – 3074
VP4U (20 mer)	5'-TCTGGCAAYACTGGTAGCAT-3'	Nucleotide sequencing	A.D.S. Bastos (pers. comm)	1314 – 1333
VP4D (20 mer)	5'-CTACCAGTATTGCCAGATTG-3'	Nucleotide sequencing	This study	1311 – 1330
SEQ1 (18 mer)	5'-CGTCGATGAGCCACTCTT-3'	Nucleotide sequencing	This study	854 – 871
SEQ2 (17 mer)	5'-CATCAAAGGCACTGAAC-3'	Nucleotide sequencing	This study	1049 – 1110
SEQ3 (18 mer)	5'-ACAACACGACACGGTACC-3'	Nucleotide sequencing	This study	1572 – 1589
SEQ4 (19 mer)	5'-TTGTGCGAAGCGTGGTTGT-3'	Nucleotide sequencing	This study	3138 – 3156
SEQ5 (18 mer)	5'-CACCAGCACGCAGTTCAA-3'	Nucleotide sequencing	This study	1859 – 1876
SEQ6 (17 mer)	5'-GTTCGTGTTCGCCAAGG-3'	Nucleotide sequencing	This study	3089 – 3105
SEQ7 (16 mer)	5'-GGTAGCAGTGGGCYGC-3'	Nucleotide sequencing	This study	3604 – 3619
SEQ8 (18 mer)	5'-ACGTTSGTCGGNGCKATG-3'	Nucleotide sequencing	This study	2134 – 2152
SEQ9 (20 mer)	5'-GCKTAACCGTAGGTGAYKCC-3'	Nucleotide sequencing	This study	1617 – 1636
SEQ10 (17 mer)	5'-GACCCBAAGACCGCAGA-3'	Nucleotide sequencing	This study	2235 – 2251
SEQ11 (18 mer)	5'-GGGAYACAGGAYTGAACT-3'	Nucleotide sequencing	This study	2630 – 2646
SEQ12 (15 mer)	5'-CGTAGATMCCCTTGT-3'	Nucleotide sequencing	This study	1785 – 1800
SEQ13 (17 mer)	5'-GCGACGTCCAACAKGTT-3'	Nucleotide sequencing	This study	2310 – 2326
SEQ14 (19 mer)	5'-CAACGGTGAGTGCAAKKAC-3'	Nucleotide sequencing	This study	3179 – 3197
SEQ15 (19 mer)	5'-GTMMTTGCACTCACCGTTG-3'	Nucleotide sequencing	This study	3179 – 3197
pUC/M13 FOR (17 mer)	5'-GTTTTCCCAGTCACGAC-3'	Nucleotide sequencing	Messing (1983)	
pUC/M13 REV (17 mer)	5'-CAGGAAACAGCTATGAC-3'	Nucleotide sequencing	Messing (1983)	

2.2.5 PCR amplification

The structural-protein-coding region (VP1-VP4) was amplified with the Expand™ Long template PCR system (Roche) using two oligonucleotides which prime in the 5'-non coding region (NCR1) and on the 2A/2B junction (WDA), respectively (Table 2.1; Figure 2.1). The reaction was carried out in the presence of 0.5 mM deoxynucleotides, 0.25 μM of each primer, 3 mM $MgCl_2$ and detergents (reaction buffer 3). After an initial denaturation step of 2 min at 95°C, thirty cycles of 30 sec denaturation at 94°C, annealing at 58°C for 30 sec and elongation for 60 sec at 68°C, were performed. To achieve a higher yield, the elongation time of each cycle was extended with 20 sec after 10 cycles and continued for the remaining 20 cycles. The resulting amplicon was analysed by agarose gel electrophoresis and recovered from the agarose gel with Cleanmix kit (Talent) according to the specification of the manufacturer.

2.2.6 T/A cloning

The purified amplicon (structural-protein-coding region) was cloned into the pGEM®-T Easy vector system (Promega) by means of T/A cloning (Figure 2.1). Ligation reactions were carried out for at least 48h at 4°C. Transformation reactions were carried out according to Tang and co-workers (1994). Plasmid isolations were prepared with the QIAprep Miniprep kit from QIAGEN.

2.2.7 Nucleotide sequencing and analysis

The nucleotide sequence of the cloned region was subsequently determined with T7 DNA polymerase (sequenase version 2.0, USB) as well as the ABI PRISM 377 DNA Sequencer from Perkin Elmer Applied Biosystems. Sequencing of the region was completed in both directions using internal primers (Figure 2.1), while analysis of the sequencing data was performed using the DAPSA version 2.9 program (Harley, 1994). Amino acid hypervariable plots were constructed with MEGA version 1.0.

2.3 Results and Discussion

2.3.1 Cloning and characterization of ZIM/7/83/2 P1

Using the oligonucleotides NCR1 and WDA, a 3kb fragment was amplified from the viral cDNA template of ZIM/7/83/2. The fragment (L-P1-2A) was subsequently cloned into the pGEM®-T Easy vector by means of T/A cloning. This vector contains 3'-T overhanging at the insertion site and greatly improves the efficiency of ligation with PCR products. The Expand long template PCR system generated PCR products have mainly 3'-A overhanging analogs and are therefore suitable for T/A cloning. Cloning and nucleotide sequencing strategies are explained in Figure 2.1. The nucleotide sequence of the cloned region is shown in Figure 2.2 together with the deduced amino acid sequence of the Leader and 2A proteinases as well as the structural proteins (VP1-4). The nucleic acid and deduced amino acid sequence data reported here for the P1 region, have been submitted to the GenBank nucleic acid sequence database (accession number AF136607).

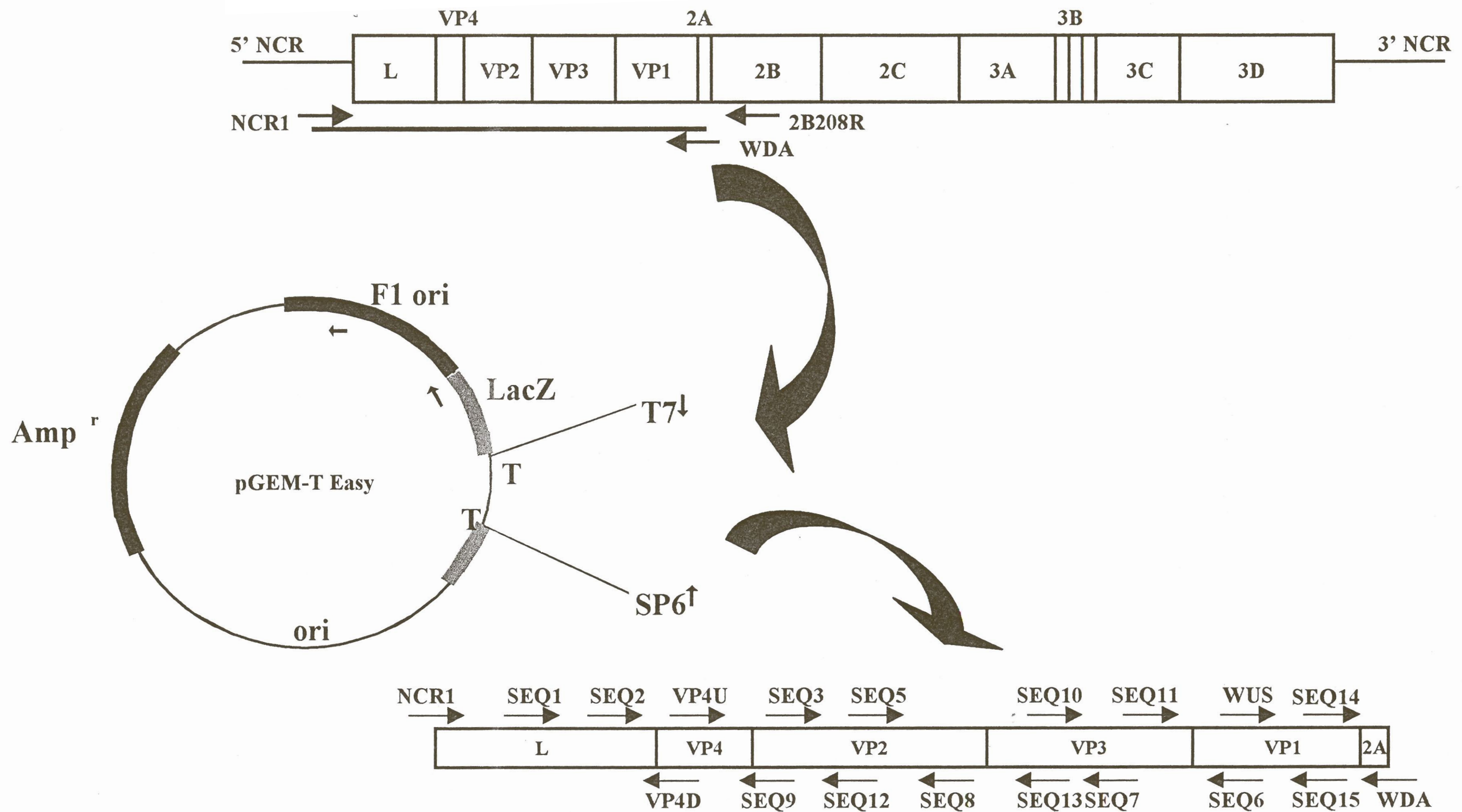

Figure 2.1: Schematic representation of the amplification of ZIM/7/83/2 P1, cloning of the amplified region into pGEM®-T Easy vector as well as the sequence strategy applied.

NCR1

1 **taacaagcgacaatcgggatct**gagaaggggaccaggagttctatcaaactgcccggtttaaaaagcttctatgcctggata

LEADER

M K T T D C F N V L

83 ggtgaccggagccggcaccttttcctttatttaaactcacttt ATG AAG ACA ACT GAC TGT TTC AAC GTT TTG

L E I I Y R F R H T F K T D R K M E F T L

156 CTC GAG ATC ATT TAC AGG TTC AGG CAC ACG TTT AAA ACA GAC AGG AAG ATG GAA TTC ACA CTC

Y N G E K K T F Y S R P N K H G N C W L N

219 TAC AAC GGA GAA AAG AAG ACC TTC TAC AGC AGG CCC AAC AAA CAC GGG AAC TGT TGG CTC AAC

S L L Q L F R Y V D E P L F E S E Y L S P

282 TCA CTT CTG CAG CTC TTT CGA TAC GTC GAT GAG CCA CTC TTT GAG TCT GAG TAC CTG TCA CCT

E N K T L D M I K Q L S D Y T K L D L S D

345 GAA AAC AAA ACA CTG GAC ATG ATC AAA CAG CTA TCT GAT TAC ACC AAA TTG GAC CTG TCA GAC

G G P P A L V L W L I K D C L Q T G V G T

408 GGA GGG CCC CCC GCT CTC GTT CTT TGG CTG ATC AAA GAT TGT CTT CAG ACT GGC GTT GGC ACC

S T R P S E I C V I N G V V M T L A D F H

471 AGC ACT CGC CCC AGC GAG ATC TGT GTC ATC AAC GGG GTT GTC ATG ACC CTG GCT GAT TTC CAC

A G I F I K G T E H A V F A L N T S E G W

534 GCC GGC ATT TTC ATC AAA GGC ACT GAA CAC GCC GTG TTC GCC CTC AAC ACA TCC GAG GGC TGG

Y A I D D E V F Y P W T P D P E N V L A Y

597 TAT GCC ATT GAT GAT GAG GTG TTC TAC CCT TGG ACA CCC GAC CCT GAA AAC GTA CTC GCG TAC

V P Y D Q E P L D V D W Q D R A G L F L R

660 GTC CCC TAC GAC CAG GAA CCA CTG GAC GTA GAC TGG CAA GAT CGC GCG GGT CTG TTC CTC CGT

VP4

G A G H S S P V T G S Q N Q S G N T G S I

723 GGA GCA GGC CAC TCA TCA CCT GTC ACA GGG TCA CAA AAC CAA TCT GGC AAT ACT GGT AGT ATC

I N N Y Y M Q Q Y Q N S M D T Q L G D N A

786 ATC AAC AAT TAC TAC ATG CAA CAG TAC CAG AAT TCA ATG GAC ACC CAA CTT GGC GAC AAC GCC

I S G G S N E G S T D T T S T H T N N T Q

849 ATC TCG GGT GGG TCC AAC GAG GGC AGC ACT GAC ACC ACG TCT ACC CAC ACA AAC AAC ACG CAG

N N D W F S K L A Q S A I S G L F G A L L

912 AAC AAT GAT TGG TTT TCA AAA TTG GCC CAG TCA GCG ATC TCG GGG CTT TTC GGA GCC CTC CTC

VP2

A D K K T E E T T L L E D R I V T T R H G

975 GCA GAC AAA AAG ACA GAG GAA ACC ACT CTG CTC GAG GAC CGC ATC GTC ACA ACA CGA CAC GGT

T T T S T T Q S S V G I T Y G Y A D A D S

1038 ACC ACC ACC TCC ACC ACA CAG AGT TCC GTT GGC ATC ACC TAC GGT TAC GCT GAC GCT GAC TCT

```
      F   R   P   G   P   N   T   S   G   L   E   T   R   V   E   Q   A   E   R   F   F
1101 TTC CGC CCC GGA CCC AAC ACA TCG GGC CTG GAG ACG CGT GTG GAA CAA GCA GAG CGG TTC TTC

      K   E   K   L   F   D   W   T   S   D   K   P   F   G   T   L   Y   V   L   E   L
1164 AAG GAA AAG CTT TTT GAT TGG ACA TCA GAC AAA CCA TTT GGC ACG CTG TAT GTT TTG GAA TTG

      P   K   D   H   K   G   I   Y   G   S   L   T   D   A   Y   T   Y   M   R   N   G
1227 CCC AAG GAC CAC AAG GGG ATC TAC GGC AGC CTG ACC GAC GCG TAT ACT TAC ATG CGC AAC GGT

      W   D   V   Q   V   S   A   T   S   T   Q   F   N   G   G   S   L   L   V   A   M
1290 TGG GAC GTC CAG GTT TCC GCC ACC AGC ACG CAG TTC AAC GGC GGG TCA CTC CTT GTG GCC ATG

      V   P   E   L   C   S   L   K   D   R   E   E   F   Q   L   S   L   Y   P   H   Q
1353 GTG CCG GAG CTG TGC TCG CTC AAG GAC AGA GAG GAG TTT CAA CTC TCT CTC TAC CCA CAC CAG

      F   I   N   P   R   T   N   T   T   A   H   I   Q   V   P   Y   L   G   V   N   R
1416 TTT ATC AAC CCA AGG ACC AAC ACC ACA GCA CAC ATC CAG GTG CCC TAC CTC GGT GTG AAC AGG

      H   D   Q   G   K   R   H   Q   A   W   S   L   V   V  ·M   V   L   T   P   L   T
1479 CAC GAT CAG GGC AAG CGC CAC CAG GCG TGG TCC CTG GTC GTC ATG GTC CTC ACG CCT CTC ACC

      T   E   A   Q   M   Q   S   G   T   V   E   V   Y   A   N   I   A   P   T   N   V
1542 ACC GAG GCA CAA ATG CAA TCC GGG ACT GTT GAG GTT TAC GCC AAC ATC GCC CCG ACG AAC GTC

                                             VP3
      F   V   A   G   E   K   P   A   K   Q   G   I   I   P   V   A   C   F   D   G   Y
1605 TTC GTT GCT GGC GAA AAG CCT GCG AAA CAG GGC ATC ATT CCA GTT GCC TGT TTC GAC GGC TAT

      G   G   F   Q   N   T   D   P   K   T   A   D   P   I   Y   G   Y   V   Y   N   P
1668 GGT GGA TTC CAA AAC ACC GAC CCG AAG ACC GCA GAT CCC ATC TAC GGT TAC GTG TAC AAC CCG

      S   R   N   D   C   H   G   R   Y   S   N   L   L   D   V   A   E   A   C   P   T
1731 TCT CGC AAC GAT TGT CAC GGC AGG TAC TCC AAC CTG TTG GAC GTC GCC GAG GCG TGC CCC ACT

      F   L   N   F   D   G   K   P   Y   V   V   T   K   N   N   G   D   K   V   M   T
1794 TTC CTG AAC TTT GAT GGT AAG CCC TAC GTC GTC ACC AAG AAC AAC GGC GAC AAG GTC ATG ACC

      C   F   D   V   A   F   T   H   K   V   H   K   N   T   F   L   A   G   L   A   D
1857 TGT TTT GAT GTG GCA TTC ACG CAC AAA GTT CAC AAG AAC ACG TTT CTT GCG GGC CTA GCG GAT

      Y   Y   A   Q   Y   Q   G   S   L   N   Y   H   F   M   Y   T   G   P   T   H   H
1920 TAC TAC GCC CAG TAC CAG GGT TCG CTG AAC TAC CAC TTC ATG TAC ACA GGT CCT ACT CAC CAT

      K   A   K   F   M   V   A   Y   I   P   P   G   I   E   T   D   R   L   P   K   T
1983 AAA GCA AAG TTC ATG GTT GCC TAC ATC CCA CCA GGC ATT GAG ACT GAC AGA CTG CCC AAG ACA

      P   E   D   A   A   H   C   Y   H   S   E   W   D   T   G   L   N   S   Q   F   T
2046 CCC GAG GAC GCA GCC CAC TGC TAC CAC TCG GAG TGG GAC ACA GGA CTG AAC TCC CAG TTC ACG

      F   A   V   P   Y   V   S   A   S   D   F   S   Y   T   H   T   D   T   P   A   M
2109 TTC GCC GTC CCA TAC GTC TCT GCA AGT GAC TTC TCC TAC ACA CAC ACT GAC ACC CCC GCA ATG

      A   T   T   N   G   W   V   A   V   F   Q   V   T   D   T   H   S   A   E   A   A
2172 GCA ACC ACC AAC GGC TGG GTG GCG GTG TTC CAG GTG ACT GAC ACC CAT TCG GCC GAA GCC GCT

      V   V   V   S   V   S   A   G   P   D   L   E   F   R   F   P   V   D   P   V   R
2235 GTG GTT GTG TCG GTG AGC GCT GGA CCC GAC CTG GAG TTC AGG TTC CCG GTT GAC CCA GTG CGC
```

VP1

Q T T S S G E G A D V V T T D P S T H G G
2298 CAA ACC ACC AGC TCA GGT GAA GGA GCG GAC GTC GTG ACG ACC GAC CCT TCG ACC CAC GGT GGT

A V T E K K R V H T D V A F V M D R F T H
2361 GCT GTC ACG GAG AAG AAA CGT GTG CAC ACA GAC GTG GCA TTC GTC ATG GAC AGA TTC ACC CAT

V L T N R T A F A V D L M D T N E K T L V
2424 GTT CTG ACA AAT AGA ACC GCG TTC GCG GTT GAC TTG ATG GAC ACC AAC GAG AAG ACC CTG GTA

G G L L R A A T Y Y F C D L E I A C L G E
2487 GGC GGC CTG CTG CGT GCG GCC ACC TAC TAT TTC TGT GAC CTG GAA ATT GCC TGC CTT GGC GAA

H E R V W W Q P N G A P R T T T L R D N P
2550 CAC GAA CGC GTG TGG TGG CAG CCA AAC GGG GCA CCG CGG ACA ACC ACG CTT CGC GAC AAC CCC

M V F S H N N V T R F A V P Y T A P H R L
2613 ATG GTG TTT TCA CAC AAC AAC GTC ACG CGT TTT GCT GTC CCG TAC ACC GCG CCA CAC CGG CTG

L S T R Y N G E C K Y T Q Q S T A I R G D
2676 CTA TCA ACC AGA TAC AAC GGT GAG TGC AAG TAC ACG CAG CAG TCC ACT GCC ATT CGC GGT GAC

R A V L A A K Y A N T K H K L P S T F N F
2739 CGT GCC GTC TTG GCC GCA AAG TAC GCC AAC ACC AAA CAC AAA CTC CCG TCT ACC TTC AAC TTC

G H V T A D K P V D V Y Y R M K R A A V Y
2802 GGC CAC GTG ACC GCC GAC AAA CCA GTC GAC GTT TAC TAC CGG ATG AAG AGG GCG GCA GTC TAC

C P R P L L P G Y D H A D R D R F D S P I
2865 TGT CCA AGA CCT CTC CTC CCT GGC TAC GAC CAC GCA GAC AGG GAC AGG TTT GAC AGC CCC ATT

2A
G V E K Q L C N F D L L K L A G D V E S N
2928 GGT GTT GAG AAA CAA CTG TGC AAC TTC GAC CTG TTG AAG TTG GCT GGA GAC GTT **GAG TCC AAC**
WDA

2B
P G P F
2991 **CCT GGG CCC TTC**

Figure 2.2: Nucleotide and deduced amino acid sequences of the LEADER and 2A proteinases as well as the structural proteins (VP1-4) of ZIM/7/83/2. The NCR1 and WDA primers used during amplification are indicated in bold.

2.3.2 Comparison of ZIM/7/83/2 P1 with known P1 regions

The deduced amino acid sequence of ZIM/7/83/2 was compared to two known SAT 2 isolates, RHO/1/48 and KEN/3/57 (Figure 2.3). Approximately 200 nucleotides on the 5'-end of VP4 of RHO/1/48 were also determined during this study. The same PCR amplification steps and T/A cloning procedures were followed as described for ZIM/7/83/2. Nucleotide sequencing of the region in question was performed with the SEQ1, SEQ2, VP4U oligonucleotides. These three viruses, all bovine outbreak strains, originate from different countries, namely Zimbabwe (ZIM/7/83), Zambia (RHO/1/48)

and Kenya (KEN/3/57), and represent different genetic lineages within the SAT 2 serotype (A.D.S. Bastos, personal communication).

Intratypic variation calculated for the individual structural proteins VP4, VP3, VP2 and VP1 of the three SAT 2 isolates amounts to 3%, 11%, 12% and 30% respectively. The VP1 region is therefore the most variable region within P1. When the intratypic variation for the P1 region and VP1 of the SAT 2 serotype is compared to that of serotypes A, O and C (Table 2.2), it is evident that the former is considerably more variable than its European counterparts. Intratypic variation for the VP1 region within the SAT 2 type (30%) is approximately two to three times higher than the variation within serotypes A, O and C.

2.3.3 Identification of hypervariable regions within the SAT 2 serotype

Field isolates of FMDV enter cells via receptor-mediated endocytosis and internalization occurs due to interaction between a protein ligand (Arg-Gly-Asp or RGD) and cell surface receptors of the integrin superfamily. Integrin receptors that have been shown to be involved in this process, include $\alpha_v\beta_3$ (Berinstein *et al.*, 1995; Jackson *et al.*, 1997; Neff *et al.*, 1998), $\alpha_5\beta_1$ (Villaverde *et al.*, 1996; Jackson *et al.*, 2000a) and $\alpha_v\beta_6$ (Jackson *et al.*, 2000b). The RGD triplet is conserved among all seven serotypes and is situated on a highly immunogenic loop in VP1. Interesting to note is the presence of an arginine residue directly after the RGD in all three SAT 2 virus sequences (Figure 2.3). This observation is in agreement with results obtained by Jackson and co-workers (2000a), who showed that this arginine residue usually follows the RGD in type SAT 2 and that these viruses have a low binding affinity for the $\alpha_5\beta_1$ receptor. In other serotypes a leucine or methionine residue usually follows the RGD and is linked to the higher binding affinity for the receptor.

The G-H loop also contains major immunodominant epitopes. Using monoclonal antibodies, these immunodominant epitopes were previously determined for RHO/1/48 (Crowther *et al.*, 1993b). One of these, valine-675 (Figure 2.3), is involved in both

```
VP4                                                                                  70
ZIM/7/83   GAGHSSPVTG SQNQSGNTGS IINNYYMQQY QNSMDTQLGD NAISGGSNEG STDTTSTHTN NTQNNDWFSK
RHO/1/48   ...Q...A.. .......... .......... .......... .......... .....F.... ..........
KEN/3/57   ...Q...A.. .......... .......... .......... .......... .......... ..........
                           VP2                                                      140
ZIM/7/83   LAQSAISGLF GALLADKKTE ETTLLEDRIV TTRHGTTTST TQSSVGITYG YADADSFRPG PNTSGLETRV
RHO/1/48   .......... .......... .......... .......... .......... .......... ..........
KEN/3/57   .......... .......... .........L .......... .......... ...S....S. ..........
                                       SITE I                                       210
ZIM/7/83   EQAERFFKEK LFDWTSDKPF GTLYVLELPK DHKGIYGSLT DAYTYMRNGW DVQVSATSTQ FNGGSLLVAM
RHO/1/48   ........K. .......... ...H...... .Q.......I ...A.T.... ....T..... ..........
KEN/3/57   .......... .......... .........R .......K.. .S........ .......... ....C.....
                                                                                    280
ZIM/7/83   VPELCSLKDR EEFQLSLYPH QFINPRTNTT AHIQVPYLGV NRHDQGKRHQ AWSLVVMVLT PLTTEAQMQS
RHO/1/48   ....S...E. .....T.... .......... .......... .......... .......... ........N.
KEN/3/57   ........A. ..Y..T.... .......... ..L....... .......... S......... .P......N.
                                     VP3                                            350
ZIM/7/83   GTVEVYANIA PTNVFVAGEK PAKQGIIPVA CFDGYGGFQN TDPKTADPIY GYVYNPSRND CHGRYSNLLD
RHO/1/48   .......... .........M .......... .A........ .......... .......... ..........
KEN/3/57   .......... ....Y....L .G....V... .A........ ....S..... .H........ ....F.....
                                                                                    420
ZIM/7/83   VAEACPTFLN FDGKPYVVTK NNGDKVMTCF DVAFTHKVHK NTFLAGLADY YAQYQGSLNY HFMYTGPTHH
RHO/1/48   .......L.. .......... .......A.. .......... ......P... .T........ ..........
KEN/3/57   .......L.D .......... .......AA. .......... ..Y....... .T..S..... ..........
                                                                                    490
ZIM/7/83   KAKFMVAYIP PGIETDRLPK TPEDAAHCYH SEWDTGLNSQ FTFAVPYVSA SDFSYTHTDT PAMATTNGWV
RHO/1/48   .......... ..V...K... .......... .......... .......... .......... .........I
KEN/3/57   ........V. ....VEE... .......... .........N .......L.S G......... ..........
                                                  VP1    SITE II                    560
ZIM/7/83   AVFQVTDTHS AEAAVVVSVS AGPDLEFRFP VDPVRQTTSS GEGADVVTTD PSTHGGAVTE KKRVHTDVAF
RHO/1/48   ..Y....... .......... .......... I........V .......... ......S.I. ...M......
KEN/3/57   V.L....... .......... .......... I........A ....E..... .T....K..T PR........
                                                                           SITE III 630
ZIM/7/83   VMDRFTHVLT NRTAFAVDLM DTNEKTLVGG LLRAATYYFC DLEIACLGEH ERVWWQPNGA PRTTTLRDNP
RHO/1/48   .L......H. SK.T.N.... ..K......A ....S..... ......V... S..F...... ....Q.G...
KEN/3/57   LL..S...H. .T...V.... ..K..A...A I..S...... ...V..V.K. KH.F...... ....Q.G...
                            SITE IV                        *         *  *           700
ZIM/7/83   MVFSHNNVTR FAVPYTAPHR LLSTRYNGEC KYTQQSTAIR GDRAVLAAKY ANTKHKLPST FNFGHVTADK
RHO/1/48   ......G.A. ..I....... ..A....... ..K.EAK... .......... .G.S.A.... ..........
KEN/3/57   ..L.R..... ..I.F..... ....V..... E..KTV.... ...E...Q.. SSA..S.... ....F.....
                   SITE V                          740 2A
ZIM/7/83   PVDVYYRMKR AAVYCPRPLL PGYDHADRDR FDSPIGVEKQ LC
RHO/1/48   A......... .EL....... .A....G... ..A....... .F
KEN/3/57   .......... .EL....A.. .A.T..GG.. ..A....A.. .L
```

Figure 2.3: Deduced amino acid sequence comparison of ZIM/7/83 (GenBank Acc. No. AF136607) with KEN/3/57 (AJ251473) and RHO/1/48 (AJ251475). The region completed in VP4 for RHO/1/48 is indicated in italics. The protein ligand RGD is boxed. Antigenic sites for RHO/1/48 are indicated with asterisks (*). Proteolytic cleavage sites between VP2/VP3, VP3/VP1 and VP1/2A are indicated in bold and italic. The myristate binding consensus sequence in VP4 is underlined. Hypervariable regions are indicated in gray shadowing.

Table 2.2: Comparison of intraypic variation between serotypes SAT 2, A, O, and C

	Serotype SAT 2[a]		Serotype A[b]		Serotype O[c]		Serotype C[d]	
	Amino acid comparison	% variation	Amino acid comparison	% variation	Amino acid comparison	% variation	Amino acid comparison	% variation
P1	[1]ZIM/7/83+RHO/1/48	8	$A_{10}+A_{12}$	6	O_1Kaufbeuren + O_1Caseros	5	Cs-8c1+ C_2Uru	4
	RHO/1/48+KEN/3/57	13	$A_{12}+A_{22}$	8	O_1Caseros + O_2 Brescia	5	Cs-8c1+C_1Germany	7
	ZIM/7/83+KEN/3/57	13	$A_{10}+A_{22}$	10	O_1Kaufbeuren + O_2 Brescia	5	C_1Germany+ C_2Uru	7
	[2]Intratypic	**16**	**Intratypic**	**11**	**Intratypic**	**7**	**Intratypic**	**9**
VP1	[1]ZIM/7/83+RHO/1/48	13	$A_{10}+A_{12}$	12	O_1Kaufbeuren + O_1Caseros	8	Cs-8c1+ C_2Uru	9
	RHO/1/48+KEN/3/57	23	$A_{12}+A_{22}$	11	O_1Kaufbeuren + O_2 Brescia	10	Cs-8c1+C_1Germany	10
	ZIM/7/83+KEN/3/57	24	$A_{10}+A_{22}$	15	O_1Caseros + O_2 Brescia	13	C_1Germany+ C_2Uru	10
	[2]Intratypic	**30**	**Intratypic**	**18**	**Intratypic**	**14**	**Intratypic**	**10**

[1]Pair-wise comparisons of isolates.
[2]Total intratypic variation within serotype.
[a]SAT 2 intratypic variation calculated for ZIM/7/83 (AF136607), RHO/1/48 (AJ251475) and KEN/3/57 (AJ251473).
[b]Serotype A intratypic variation calculated for A_{10}(X00429), A_{12} (M10975) and A_{22} (X74812).
[c]Serotype O intratypic variation calculated for O_1Kaufbeuren (X00871), O_1 Caseros (U82271) and O_2Brescia (M55287).
[d]Serotype C intratypic variation calculated for Cs-8c1 (M60118), C_1Germany/26 (M90368), C_2Uruguay/44 (M90367).

conformational and non-conformational epitopes and is conserved in all three strains. In contrast, an immunodominant site present at residues 682 and 684 was found to be highly variable within the SAT 2 serotype.

Using an amino acid hypervariable plot constructed with MEGA 1.0, five independent hypervariable regions were determined. The region in VP2 (167–189) corresponds with immunodominant site 2 of serotype O (Kitson *et al.*, 1990) as well as immunodominant site D of serotype C (Mateu *et al.*, 1990). The second hypervariable region (533-581) corresponds with immunodominant site 3 of serotype O (Kitson *et al.*, 1990). Residues 647 to 697 contain the G-H loop, while residues 709 to 731 contain the C-terminus of VP1. Both regions have previously been shown to contain major immunodominant sites for the other serotypes (Mateu, 1995). This is however only an indication of hypervariability within the SAT 2 serotype and the actual immunodominant sites need to be determined empirically, e.g. through escape mutant studies.

2.3.4 Description of ZIM/7/83/2 cleavage sites

Although the 3C proteinase cleavage sites for the FMDV P1 region will be discussed in more detail in Chapter 3, it is interesting to note some differences and similarities between types A and SAT 2. The P1 region is cleaved by the 3C proteinase to produce VP0 (E/G), VP3 (Q/T) (Belsham, 1993) and VP1 (Q/) (N. Knowles, personal communication), according to the A_{12} sequence (Robertson *et al.*, 1995; Baxt *et al.*, 1989). The cleavage site between VP0 and VP3 of A_{12} contains an acidic glutamic acid (E), while a polar glutamine (Q) residue is found instead in the SAT 2 serotype. The VP3/VP1 as well as the VP1/2A junction sites of the SAT 2 serotype corresponds with that of A_{12}. It is interesting to note that the flanking regions, which could play a role in the recognition of the cleavage site by the proteinase, are not highly conserved within the SAT 2 serotype. These differences could alter the structure of the cleavage site and have implications in the construction of recombinant viral particles considering that a 3C proteinase from a particular serotype may not be able to optimally process the structural proteins of a different viral serotype.

2.4 Conclusions

As has been shown previously and verified again in this study, the VP1 region is the most variable within the P1 region (Palmenberg, 1989). Therefore, an alternative option in the construction of chimeric viruses could be to exchange the VP1 region only. Advantages would be the exclusion of proteolytic cleavage sites, as well as the simplification of the PCR-cloning steps due to the smaller size of the domain to be exchanged. Retaining all the other structural proteins may also decrease the likelihood of the recombinant particle being less stable than the wild type vaccine isolate. However, it is known that both conformational and non-conformational epitopes exist for FMDV (Crowther *et al.*, 1993a; Mateu, 1995). Thus, the expected immune response may not be achieved if only VP1 is exchanged, due to altering of immunodominant sites upon assembly of the virion. Towards the construction of infectious chimeric viruses, it is imperative to investigate the heterogeneity in the FMDV proteinases and thus to understand the role of proteolytic variation and cleavage site differences in isolates originating from different geographical localities.