Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Available online at www.sciencedirect.com
ScienceDirect
Virology 357 (2007) 124–133
VIROLOGY
www.elsevier.com/locate/yviro

Synthesis, proteolytic processing and complex formation of N-terminally nested precursor proteins of the Rift Valley fever virus glycoproteins

Sonja R. Gerrard [a,b,c,1], Stuart T. Nichol [a,*]

[a] Special Pathogens Branch, Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, 1600 Clifton Rd., NE, MS G-14, Atlanta, GA 30333, USA
[b] Department of Epidemiology, School of Public Health, University of Michigan, 109 S. Observatory, Ann Arbor, MI 48109, USA
[c] Department of Microbiology and Immunology, School of Medicine, University of Michigan, 109 S. Observatory, Ann Arbor, MI 48109, USA

Received 14 April 2006; returned to author for revision 24 July 2006; accepted 2 August 2006
Available online 8 September 2006

Abstract

The genomic M RNA segment of Rift Valley fever virus is transcribed to produce a single mRNA with multiple translation initiation sites. The products of translation are an N-terminal nested series of polyproteins. These polyproteins enter the secretory system of the host cell and are proteolytically processed to yield the mature virion glycoproteins, Gn and Gc, and two non-structural glycoproteins. By means of pulse-chase immune precipitation experiments we identify the Gn and Gc precursor molecules and also show that signal peptidase cleavage is required for mature Gn and Gc production. We also demonstrate that a hydrophobic domain at the N-terminus of Gn acts as a signal peptide only in the context of the polyprotein precursors that initiate at the second, fourth or fifth AUGs. In addition, we document that formation of Gn/Gc heteromeric complexes occur rapidly (<5 min) and can occur prior to signal peptidase processing of Gn, suggesting that this complex forms in the endoplasmic reticulum. Interestingly, Gc can form a complex with a glycoprotein that has been considered nonstructural, a discovery that has implications for both the topology and potential packaging of this glycoprotein.
Published by Elsevier Inc.

Keywords: Rift Valley fever virus; Glycoproteins

Introduction

The *Bunyaviridae* family is comprised of a diverse collection of segmented negative and ambisense RNA viruses, many of which are known pathogens of humans, livestock or plants. Examples of these are Rift Valley fever (RVF) virus, Crimean-Congo hemorrhagic fever virus, hantaviruses, tomato spotted wilt virus and several viruses that cause encephalitis (Nichol, 2001). RVF virus is of particular interest because of the severity of the disease in humans and livestock and because of its demonstrated ability to spread from its traditional confines of sub-Saharan Africa. RVF is an arthropod and aerosol-borne disease that affects a wide variety of animals, including humans, sheep, cows, goats, and indigenous wild animals (Easterday, 1965; Meegan and Bailey, 1988–1989). RVF virus has caused devastating epidemics in sub-Saharan Africa. In the 1997–98 epidemic in Kenya, Somalia, and Tanzania and it was estimated that 14% of the human population in the Garissa District of Kenya was infected with the virus (Nichol, 2001; Woods et al., 2002).

Livestock are more severely affected by the disease than humans, causing 'abortion storms' and mortality in sheep approaching 100% for newborn and up to 60% for adults (Nichol, 2001). The broad vertebrate host range coupled with the fact that many different mosquito species can transmit the virus (Nichol, 2001; Easterday, 1965; Jupp et al., 2002; Turell and Kay, 1998; Turell et al., 1996; Jupp and Cornel, 1988; Gargan et al., 1988; Gad et al., 1987), results in a high probability that RVF virus will impact regions outside of its traditional confines of sub-Saharan Africa in the future. In the year 2000, there was a RVF epidemic in Saudi Arabia and

* Corresponding author. Fax: +1 404 639 1118.
E-mail addresses: gerrard@umich.edu (S.R. Gerrard), stn1@cdc.gov (S.T. Nichol).
[1] Fax: +1 734 764 3192.

0042-6822/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.virol.2006.08.002

Yemen and this epidemic represented the first time RVF activity was documented outside the African continent (Anonymous, 2000; Shoemaker et al., 2002). Human disease ranges from uncomplicated febrile illness to more severe manifestations with high case fatality, including hemorrhagic fever and encephalitis (Nichol, 2001). There are no licensed vaccines in the U.S. currently available for RVF virus and no effective therapeutic drugs. As the glycoproteins, Gn and Gc, contain the only source of neutralizing epitopes, it is important to characterize their processing and maturation for use in future vaccine development.

RVF virus belongs to the genus *Phlebovirus* within the family *Bunyaviridae* and like all RNA viruses, has an extremely small genome. The two negative-stranded and one ambisense RNA genomic segments (L, M, S) combined are 12 kB in size and half of the genome (the L segment) is dedicated to encoding the RNA-dependent RNA polymerase. Given the constraints placed upon the coding potential by the small genome size, it is not surprising that RNA viruses use many unique RNA editing and translational methods to increase the repertoire of proteins synthesized from their genomes (Holland and Domingo, 1998; Moya et al., 2004). The RVF virus M segment, which encodes the envelope glycoproteins, Gn and Gc, has at least 4 translation initiation sites within the single mRNA transcribed from the genomic M segment. The translational products of the M segment are thus a nested set of polyproteins (Suzich et al., 1990; Collett, 1986). These polyproteins enter the secretory system of the host cell where they are processed by cellular enzymes into mature Gn and Gc as well as at least two additional proteins. Amongst the other molecularly characterized members of the genus *Phlebovirus*, there is considerable variability, however the Gn and Gc envelope glycoproteins still share obvious structural features such as spacing of cysteine residues and transmembrane domains (Gro et al., 1997). However, this homology does not extend to the highly variable N-terminus of the polyproteins. Initiation at the first or second AUG of the RVF virus M segment results in production of proteins that have been designated NSm, based on the presumption that they are non-structural. Although the function of these proteins has not yet been elucidated, initial investigation suggested that they are not required for proteolytic processing or intracellular transport of the mature Gn and Gc envelope glycoproteins (Wasmoen et al., 1988; Suzich and Collett, 1988). The viruses of the family *Bunyaviridae* also share another important feature and that is maturation in the Golgi apparatus (Schmaljohn and Hooper, 2001). It is thought that the envelope glycoproteins alone contain Golgi localization information and thus the glycoproteins recruit the remaining structural elements to the Golgi prior to budding (Schmaljohn and Hooper, 2001; Gerrard and Nichol, 2002). It has been shown for several members of the genus *Phlebovirus* (e.g. Punta Toro virus (PTV) (Chen and Compans, 1991), and Uukuniemi (UUK) virus (Persson and Pettersson, 1991)) that Gn and Gc can be isolated as a heteromeric complex from mature virions. It is thought that oligomerization of the viral glycoproteins is critical for their proper transit to the Golgi apparatus as only Gn contains a Golgi localization signal (Pettersson et al., 1995; Melin et al., 1995; Shi et al., 2004; Spiropoulou et al., 2003; Chen and Compans, 1991; Gerrard and Nichol, 2002).

In this report, we present a detailed characterization of the RVF virus translational products of the virus M segment and also show that signal peptidase cleavage at two sites is required for mature Gn and Gc production. Surprisingly, the hydrophobic domain at the N-terminus of Gn acts as a signal peptide only in the polyprotein precursors that initiate at the second, fourth or fifth AUG. When translation of the M segment mRNA initiates at the first AUG this hydrophobic region does not function as a signal peptide. We document that formation of the Gn and Gc heteromeric complex occurs rapidly (<5 min) and does not require signal peptidase processing of Gn, suggesting that this complex forms in the endoplasmic reticulum (ER). Interestingly, Gc can form a complex with one of the NSm proteins, a discovery that has implications for both the topology and potential packaging of this presumed non-structural protein.

Results

Polyprotein sequence features

The M genomic segment of RVF virus encodes the two structural glycoproteins, Gn and Gc, as well as at least two other proteins both of which have been referred to as NSm (*n*on-*s*tructural protein encoded by the *M* segment) (Fig. 1A) (Suzich et al., 1990). However, it is possible that one or both of these NSm proteins are present in the virion at low abundance, much like the E protein of coronaviruses (Godet et al., 1992). The mRNA for the M segment contains 5 in-frame AUG codons within the NSm coding region (Fig. 1B), and it has been shown that 4 of these AUG codons are utilized in the synthesis of an N-terminal nested set of polyprotein precursors (Suzich et al., 1990). Interestingly, only one of these AUG codons, the fourth, is in the context of an optimal translational start context sequence (Kozak, 2002), yet translation initiates at each AUG codon in this region, with the exception of the third (Suzich et al., 1990). We will refer to the proteins containing the NSm region by the AUG codon from which they initiate, for example NSm1, is the mature 68 kDa protein that initiates at the first AUG and NSm2-Gn is a 64 kDa precursor protein to Gn initiated at the second AUG (Fig. 1A). In the case of NSm1, the protein begins with what appears to be a signal peptide or signal anchor (Fig. 1A). The C-terminus of NSm1 is identical to that of Gn (Fig. 1A) (Kakach et al., 1988). NSm2-Gn does not have an N-terminal signal peptide yet enters the secretory pathway by virtue of the signal peptide upstream of Gn (Fig. 2A). Polyproteins that initiate at either the fourth or fifth AUG produce only Gn and Gc glycoproteins (Fig. 2B).

There are three potential signal peptidase cleavage sites that precede NSm1, Gn and Gc, respectively (Fig. 1A) (Nielsen et al., 1997). The predicted signal peptidase cleavage sites are in agreement with the experimentally determined amino termini of Gn and Gc (Collett et al., 1985). The precise amino terminus of

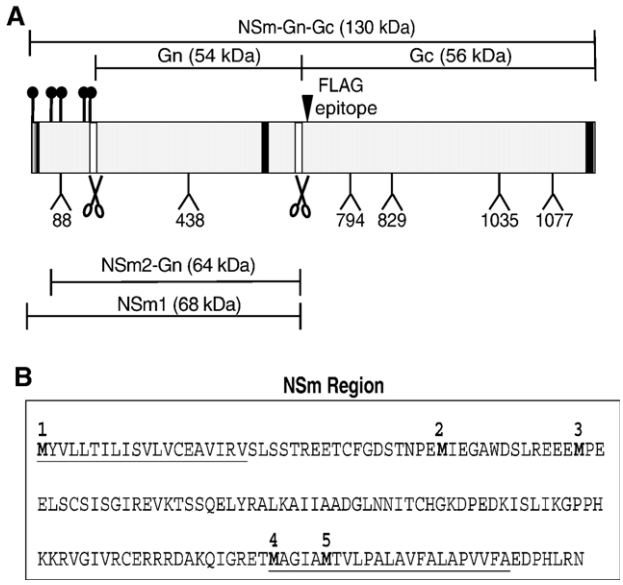

Fig. 1. (A) Structural features of the RVF virus M segment RNA and encoded proteins. Scissors symbols indicate putative signal peptidase processing sites, ball and stick symbols indicate the position of the in-frame start codons in the NSm region, and the branch and stick symbols indicate the position of putative N-linked glycosylation sites. The glycosylation site numbering system refers to the amino acid position and is relative to translation initiation at the first start codon. The brackets delineate the mature and precursor proteins along with their calculated molecular weights. The black, white and hatched boxes indicate transmembrane, signal peptide and uncharacterized hydrophobic domains, respectively. (B) The NSm region of the polyprotein precursor molecule. The translation of the NSm region is in single letter code. The hydrophobic domain at the amino terminus of NSm1 and the signal peptide upstream of Gn are underlined and the methionine (M) residues are bolded and numbered.

NSm1 has not been experimentally determined, and we do not address the issue of signal peptidase cleavage at this site within this report. Potential topologies for the polyprotein precursor molecules are diagrammed in Fig. 2. The topology of NSm1 is not known, however there is one N-linked glycosylation site at amino acid 88 and one within the Gn region, at amino acid 438, and both sites have been reported to be utilized (Fig. 1A) (Kakach et al., 1989). There are several potential explanations for this surprising observation and we address this issue in Discussion. The topology of the structural glycoproteins, Gn and Gc, is known from both mapping of neutralizing epitopes and also N-linked glycosylation usage (Kakach et al., 1989; Keegan and Collett, 1986) (Fig. 2B).

Context is critical for determining if a hydrophobic domain is utilized as a signal anchor or a signal peptide (Nielsen and Krogh, 1998; Bendtsen et al., 2004; Nielsen et al., 1997). Signal peptides differ from signal anchors in many ways, including length of the hydrophobic domain, length of the amino terminal region upstream of the hydrophobic domain and specific amino acid content within the hydrophobic and flanking domains (Nielsen and Krogh, 1998; Nielsen et al., 1997). One method for predicting signal peptides is a hidden Markov model that compares an input protein to a database of known signal peptides and signal anchors (Nielsen and Krogh, 1998). This model evaluates hydrophobic regions for potential as signal peptides by comparing the amino terminal region (n-region), the hydrophobic region (h-region) and signal peptidase cleavage region (c-region) against a set of known signal peptides and signal anchors. A probability can then be assigned to a sequence acting as a signal peptide based on the evaluation of n-, h- and c-regions. A cleavage probability score of 1.0 indicates that the sequence is strongly predicted to be a signal peptide. When translation initiates from the fourth AUG, the N-terminus of the polyprotein is strongly predicted to be a signal peptide with a cleavage probability of 1.0 (Fig. 2D), the same result is obtained when translation initiates at the fifth AUG (data not shown). However, when translation initiates from the second AUG, the hydrophobic domain upstream of Gn is not predicted to be an efficient signal peptide (Fig. 2C). The probability for the hydrophobic domain upstream of Gc is 0.702, and is thus predicted to represent a signal peptide (data not shown).

Identification of precursor and mature glycoproteins

In an attempt to identify the precursor molecules to the structural glycoproteins, Gn and Gc, we performed pulse-chase immune precipitations. To accurately determine which proteins began with NSm encoded sequences, we utilized two constructs, one which had the entire coding sequence of the native M segment (RVFMGcFLAG) and one which began at the fourth AUG (ΔNSmGcFLAG), thus eliminating the NSm region (Fig. 1A). All constructs contained a FLAG epitope after the putative signal peptidase cleavage site within Gc (Fig. 1A). We used as controls expression constructs which contained only Gn (untagged) or Gc (with the FLAG epitope in the same position as the full-length constructs) sequences. The FLAG epitope serves two purposes, it increases the molecular weight of Gc, which allows for better resolution of Gc and Gn on gels, and also allows Gc to be immune precipitated regardless of conformation, as our Gc specific monoclonal antibodies appear to recognize only mature, Golgi localized, Gc (data not shown).

Cells expressing our precursor constructs were pulsed with ^{35}S-cysteine for 5 min and then chased for the times indicated on the figures. Proteins were then immune precipitated with antibodies that recognize either Gn or FLAG. The banding pattern of our immune precipitated proteins is complex and reflects the fact that we are observing precursor processing, Gn and Gc complex formation and possibly also glycoprotein precursors interacting with cellular proteins during the folding process (Fig. 3). However, by comparing the banding given by expression of the various glycoprotein constructs it is possible to identify a number of precursor molecules. Beginning with the high molecular weight bands, we can identify at least two major precursor molecules, labeled NSm-Gn-Gc or Gn-Gc (Fig. 3). NSm-Gn-Gc (i.e. the protein initiating at the first AUG) is the more abundant precursor, and the increased size relative to Gn-Gc reflects the addition of the NSm region. This is in agreement with previously published studies that convincingly show the usage of at least 4 different initiator methionine codons in glycoprotein synthesis (Suzich et al., 1990).

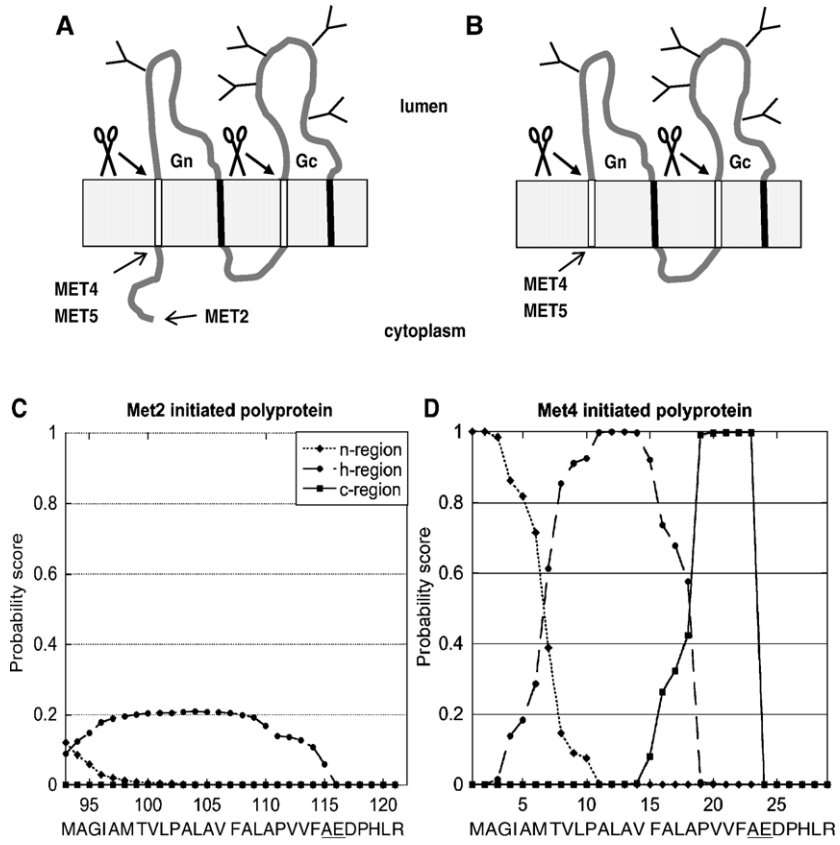

Fig. 2. Panels A and B, potential topologies of the polyprotein precursors that initiate at the second AUG (Panel A), or fourth or fifth AUG (Panel B). Signal peptidase cleavage sites are indicated by scissors, putative N-linked glycosylation sites are indicated by branch and stick symbols, signal peptides are the white boxes and transmembrane domains are black boxes. Locations of initiator methionines (MET) are indicated. Panels C and D, hidden Markov model predictions for the hydrophobic domain preceding Gn in the context of the polyproteins beginning at MET2 (Panel C), or MET4 (Panel D). The n-, h- and c-regions refer to the domain proximal to the hydrophobic region, the hydrophobic region and the signal peptidase cleavage region, respectively. The probability that the sequence contains an n-, h- or c- region of a signal peptide is plotted on the *y*-axis, and the amino acids and their positions are plotted on the x-axis. The plots encompass the hydrophobic domain and begin at (C) amino acid 93 for MET2 initiated polyprotein and at (D) amino acid 1 for the MET4 polyprotein. The signal peptidase cleavage site is between the alanine and glutamic acid residues that are underlined.

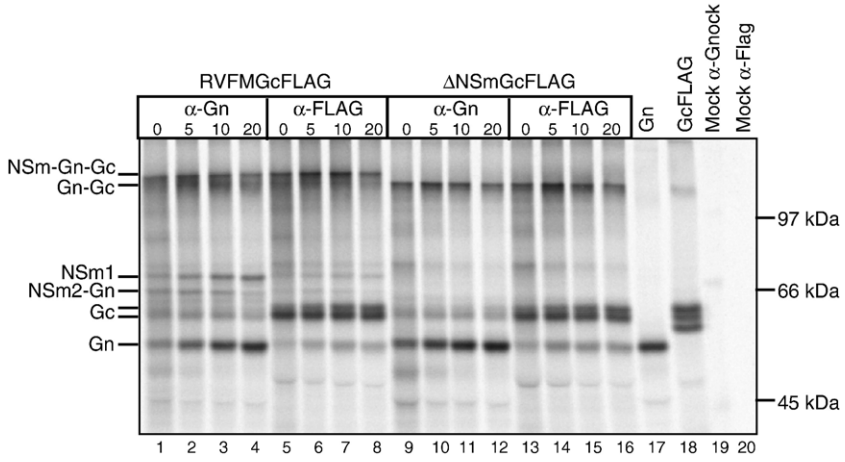

Fig. 3. Multiple polyprotein precursor molecules are generated from RVF virus M segment mRNA. BSR-T7/5 cells transformed with either the RVFMGcFLAG or ΔNSmGcFLAG plasmid were pulse labeled for 5 min with ^{35}S-cysteine, then chased for the times, expressed in minutes, indicated at the top of the gel. Immune precipitation of M segment encoded glycoproteins, with antibodies recognizing either Gn or FLAG, was carried out as described in Materials and methods. The lane numbers are indicated at the bottom of the gel.

There are two additional NSm encoded proteins; one at 68 kDa, labeled NSm1, and another at 66 kDa, labeled NSm2-Gn (compare lane 1 with lane 9 in Fig. 3). NSm1 is the product of initiation at the first AUG and although it contains the Gn signal peptidase cleavage site (Fig. 1A), it is not proteolytically processed within the 20 min chase period (lanes 1–4, Fig. 3) nor is it processed after chase times as long as 40 min (data not shown). NSm2-Gn initiates at the second AUG and appears to be slowly processed during the 20 min of chase time (lanes 1–4, Fig. 3) (Suzich et al., 1990). Evidence that NSm2-Gn is further processed to generate Gn is discussed below in the section dealing with signal peptidase usage (Fig. 4). Gc runs as a doublet at 59 kDa when expressed from the authentic precursor molecule (Fig. 3, lane 8) and as a triplet when expressed alone (lane 18, Fig. 3). Gc is glycosylated (Kakach et al., 1989) and is predicted to be palmitoylated, thus it is possible that the doublet is a result of differential post-translational modification. A Gc doublet is also seen under similar experimental conditions when RVF virus-infected cells are examined (data not shown), therefore we believe that the doublet is not an artifact of epitope tagging the protein or of our plasmid-based expression system. When Gc is expressed in the absence of Gn (as in lane 18, Fig. 3), it localizes to the ER, whereas when expressed from the polyprotein precursor (as in lane 8, Fig. 3), it localizes to the Golgi (Gerrard and Nichol, 2002). Thus, it is likely that the extra band in the lane where Gc is expressed by itself (lane 18, Fig. 3) is an aberrant form of the protein that is a direct result of its mislocalization. Gn expressed from a polyprotein precursor runs as a single band at 54 kDa (lane 4, Fig. 4), and is indistinguishable from Gn expressed alone (lane 17, Fig. 3).

There are several minor bands visible in the 70–90 kDa range, which are observed at the earliest chase times, are not affected by presence or absence of NSm encoded sequences or by mutation of signal peptide cleavage sites (Figs. 3–5), and not found when immune precipitations are performed on mock transfected cells (lanes 19–20, Fig. 3). It has been shown that the glycoproteins from UUK virus, another member of the genus *Phlebovirus*, interact with the ER chaperonins calnexin (88 kDa) and calreticulin (60 kDa) (Veijola and Pettersson, 1999). While we cannot definitively say that these proteins are cellular in origin, we feel that given the timing of their appearance and disappearance it is likely that these are indeed ER chaperonins.

Gn and Gc interaction

Although the immune precipitations were performed in the presence of ionic and non-ionic detergents, we found that Gn and Gc complexes can still be detected (Fig. 3). Gc and Gn co-immune precipitate at the earliest chase time, and the amount of Gc co-immune precipitating with Gn does not increase over the time frame of the chase (lanes 1–4, Fig. 3). This result suggests that Gn and Gc form complexes extremely quickly, within the 5 min of our radioactive pulse, and that complexes are forming between newly synthesized Gn and Gc. Furthermore, given the rapid kinetics, complex formation likely occurs in the ER. This is consistent with data for most members of the family *Bunyaviridae* where Gn and Gc generally form multimers and where complex formation is essential for proper Gc localization; Gc localizes to the ER when expressed alone, but localizes properly to the Golgi apparatus when co-expressed with Gn (Schmaljohn and Hooper, 2001). Here, rather surprisingly we show that NSm1 and NSm2-Gn also co-immune precipitate with Gc (lanes 5–8, Fig. 3). NSm2-Gn diminishes over the chase, suggesting that although it is in a complex with Gc, this complex does not abrogate processing by signal peptidase (lanes 5–8, Fig. 3). The fact that Gc can form a complex with NSm2-Gn likely explains why the amount of Gn co-immune precipitating with Gc increases over the chase period, while the amount of Gc co-immune precipitating with Gn remains constant over the chase period (lanes 1–8, Fig. 3). NSm1 is essentially a fusion protein between the NSm region and the Gn envelope glycoprotein (Fig. 1A). The exact topology of the NSm1 is ambiguous because the N-linked glycosylation sites at

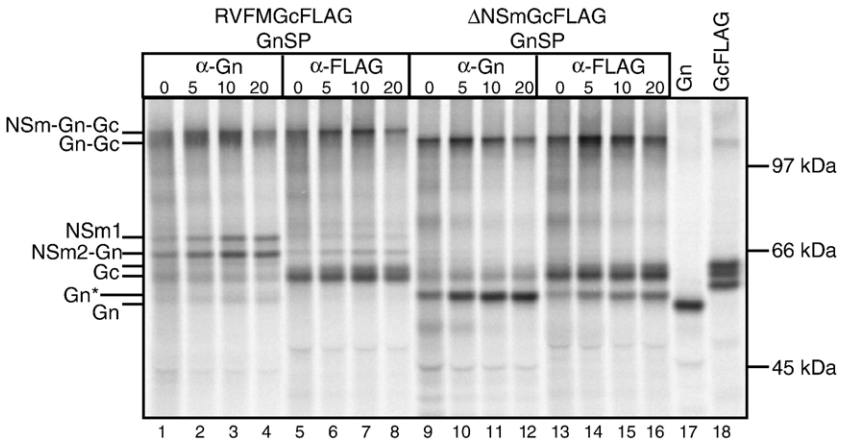

Fig. 4. Signal peptidase cleavage at the site within Gn is required for generation of mature Gn. BSR-T7/5 cells transformed with either the RVFMGcFLAG-GnSP or ΔNSmGcFLAG-GnSP plasmid, both plasmids contain an alanine to arginine mutation at the predicted signal peptidase cleavage site within Gn. Cells were pulse labeled for 5 min with ^{35}S-cysteine, then chased for the times, expressed in minutes, indicated at the top of the gel. Immune precipitation of M segment encoded glycoproteins, with antibodies recognizing either Gn or FLAG, was carried out as described in Materials and methods. The lane numbers are indicated at the bottom of the gel.

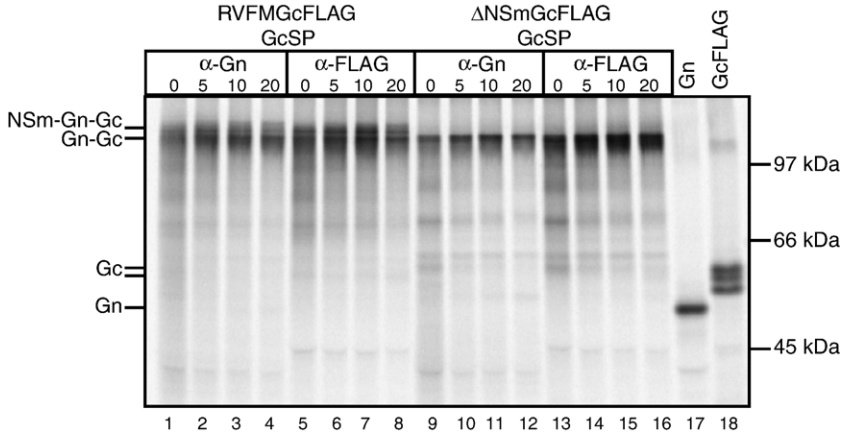

Fig. 5. Signal peptidase cleavage at the site within Gc is required for generation of all mature RVF virus M segment encoded glycoproteins. BSR-T7/5 cells transformed with either the RVFMGcFLAG-GcSP or ΔNSmGcFLAG-GcSP, both plasmids contain an alanine to arginine mutation at the predicted signal peptidase cleavage site within Gc. Cells were pulse labeled for 5 min with ^{35}S-cysteine, then chased for the times, expressed in minutes, indicated at the top of the gel. Immune precipitation of M segment encoded glycoproteins, with antibodies recognizing either Gn or FLAG, was carried out as described in Materials and methods. The lane numbers are indicated at the bottom of the gel.

position 88 and 438 of NSm1 have been reported to be utilized (Fig. 1A) (Kakach et al., 1989). It is known that Gn contains a Golgi localization signal within a region that includes the transmembrane domain and a short segment within the cytosolic tail (Gerrard and Nichol, 2002), and that NSm1 also localizes to the Golgi apparatus (Wasmoen et al., 1988). Gc is thus, either interacting with NSm1 in the same manner (and membrane orientation) as with Gn, or it is forming a novel complex with the N-terminal portion of NSm1 (Fig. 1A).

Gn and Gc are cleaved from the precursor by signal peptidase

In order to test the hypothesis that the cellular protein, signal peptidase, is required for polyprotein processing, we mutated the putative Gn signal peptidase cleavage site within our RVFM and ΔNSm constructs to generate RVFMGnSP and ΔNSmGnSP, respectively. The alanine→arginine mutation at the −1 position of the predicted cleavage site abolished Gn accumulation consistent with the hypothesis that signal peptidase cleavage is required for polyprotein processing (compare Fig. 4 with Fig. 3) (Nielsen et al., 1997). Additionally, we found that a new protein appeared that ran slightly higher than Gn, which we have designated Gn* (Fig. 4). The mobility of Gn* is consistent with it being Gn with the signal peptide still attached. NSm2-Gn, was stabilized as a result of this mutation, in support of our hypothesis that it is a precursor to Gn (Fig. 4). Interestingly, the rate of processing for the Gn signal peptide in the context of NSm2-Gn is much slower than when it is in the context of the Gn-Gc polyprotein (compare lanes 1–4 in Fig. 3). Gn* is only apparent when the molecule is stabilized by mutation of the signal peptide cleavage site, yet NSm2-Gn is clearly visible at the 20 min chase time. This result is in agreement with the signal peptide predictions shown in Fig. 2, and may be a result of inefficient insertion of the NSm2-Gn precursor into the ER. NSm1 and NSm2-Gn both contain the Gn signal peptidase cleavage site, yet only NSm2-Gn is a precursor to Gn. Why NSm1 is not cleaved to liberate Gn is not clear but may indicate that these proteins have different topologies. More Gn* is made with the ΔNSmGnSP construct than with the RVFMGnSP. This is probably attributable to there being are at least three different mechanisms to generate Gn when the NSm sequences are present, the precursors initiating at the second, fourth and fifth AUG. When synthesis begins at the second AUG, NSm2-Gn results, while synthesis beginning at the fourth or fifth AUG results in Gn*. Gc accumulated in a manner indistinguishable from our wild-type constructs and co-immune precipitated with Gn* and NSm2-Gn which is consistent with our data presented in Fig. 3 and further supports our hypothesis that Gn and Gc complexes form in the ER (Fig. 4).

Analogous mutations to the Gn signal peptidase cleavage sites were made in Gc, and the plasmids designated RVFMGcSP and ΔNSmGcSP. No mature proteins accumulated when the Gc signal peptidase cleavage site was mutated and the high molecular weight precursors NSm-Gn-Gc and Gn-Gc were stabilized (Fig. 5). Therefore, signal peptidase cleavage at Gc is required for generation of all mature proteins synthesized from the RVF virus M segment. This result does not rule out processing by other cellular endoproteases, however it does indicate that signal peptidase cleavage is a prerequisite for any further processing of the polyproteins.

Discussion

Using pulse-chase immune precipitation, we have shown that RVF virus glycoproteins are synthesized from several precursor polyproteins that differ only at the N-terminus and that the cellular protein, signal peptidase, is required for proteolytic processing of the precursor molecules into mature envelope glycoproteins and proteins containing the NSm region. We have also shown that newly synthesized Gc rapidly

forms complexes with Gn, Gn precursors and NSm1. The kinetics of complex formation suggests that these events occur within the ER.

It is not yet known if other viruses of the genus *Phlebovirus* use multiple start codons in the synthesis of their non-structural and structural glycoproteins. Analysis of the 4 other members of the genus reveals no amino acid sequence identity within the NSm region of the polyprotein, and in fact one member of the genus, UUK virus, encodes no NSm proteins. However, the NSm region of PTV, SSF virus and Toscana virus do share some similarities. The first AUG appears to initiate translation at a predicted signal peptide and the sequence between this signal peptide and the Gn signal peptide is highly charged and acidic (pI<5.5), sequence properties similar to those seen in the NSm region of RVF virus. Eukaryotic translation has been shown to initiate efficiently when the AUG is in the context of RNNaugG (Kozak, 2002). When the first AUG is not in this context, the ribosome will often continue scanning in the 3′ direction until an AUG in the preferred context is located (Kozak, 2002). The NSm region of PTV and SSF virus appear to share similar coding strategy to that of RVF virus in that there are multiple in-frame AUG codons, the first AUG is not in the context of a strong initiator sequence, and at least one of the downstream AUG codons is in the proper context. By contrast, UUK virus M segment mRNA lacks the NSm coding region and the first AUG is within an efficient initiation context (data not shown). Whether these multiple initiator codons are actually utilized is only known for RVF virus. However, the NSm sequences within PTV, SSF virus and Toscana virus also appear to all be designed to promote 'leaky scanning' by the host translational machinery and thus generate nested sets of polyproteins differing only at the N-terminus. The significance of the NSm region for these viruses remains to be determined. However, given that viruses of the genus *Phlebovirus* only have 4 genes and evolutionary pressure exists to reduce RNA virus genome size, it is likely that this region is required during some aspect of their natural transmission cycle. It should also be noted that while these phleboviruses all infect humans and other vertebrates, they use different arthropod vectors (RVF virus; mosquitoes, PTV; New World sandflies, SSF and Toscana virus; Old World sandflies, UUK virus; ticks) and it is possible that the NSm region plays some essential role in arthropod infection.

The hydrophobic domain preceding Gn is a signal peptide only when translation initiates at the second, fourth or fifth AUG. Signal peptidase processes the Gn signal peptide in the context of the Gn-Gc polyprotein rapidly, whereas this same cleavage site is processed very slowly in the context of NSm2-Gn. Mutation of the cleavage site preceding Gn stabilizes Gn* and NSm2-Gn, therefore both proteins are processed by signal peptidase. The precise cause of this rate differential is not known, however, there is not an N-terminal signal peptide on this precursor protein and the hydrophobic domain upstream of Gn is not predicted to act as an efficient signal peptide (Fig. 2), thus the mode of insertion into the ER membrane may be quite different from that of the Gn-Gc polyprotein. The data with respect to the topology of NSm1 is conflicting, as N-linked glycosylation sites on either side of a transmembrane domain are utilized (Fig. 1A) (Kakach et al., 1989) and the signal peptidase cleavage site preceding Gn is not utilized (Fig. 3). Since the glycosylation enzymes and signal peptidase are in the lumen of the ER, it is not clear how NSm1 is inserted into the membrane. It is possible that the orientation of the NSm1 protein is altered post-translationally, although this would involve changing the orientation of at least two transmembrane domains. Alternatively, the hydrophobic domain preceding Gn may not be a transmembrane domain when it is in the context of NSm1. Experiments to further characterize the processing and topology of NSm1 are underway.

Gn and Gc carry conflicting intracellular sorting information within their primary sequence. Gn contains a Golgi localization motif and Gc an ER retrieval motif, yet both proteins localize to the Golgi in steady-state when co-expressed (Gerrard and Nichol, 2002). This pattern is consistent throughout the family *Bunyaviridae* and suggests that while these viruses have diverged significantly over the course of their evolution, the basic mechanism of virus assembly has remained constant. The fact that both mature structural glycoproteins localize to the Golgi apparatus, despite their conflicting intracellular sorting signals suggests that Gn and Gc form a complex. Complexes of Gn and Gc have been clearly documented for UUK virus (Persson and Pettersson, 1991), and PTV (Chen and Compans, 1991). In addition, the conservation of a canonical dilysine based ER retrieval motif within the Gc of 3 genera within the family *Bunyaviridae* suggests that localization of Gc to the ER is extremely important for the maturation process (data not shown). These observations together point towards the ER as the likely place for Gn and Gc complex formation, therefore finding Gc and Gn co-immune precipitating within the first 5 min after synthesis is not surprising. What is surprising is finding NSm1 co-immune precipitating with Gc. NSm1 is thought to be non-structural, however it is a fusion of the NSm region with Gn and it localizes to the Golgi apparatus, the site of viral maturation (Fig. 1A) (Wasmoen et al., 1988). NSm1 is synthesized at much lower levels than Gn, therefore it is possible that NSm1 is packaged and is a minor component of the virion. Whether or not this protein is packaged into RVF virions remains to be determined, as does its role in maintenance of RVF virus in nature. Establishing the functional role of this novel fusion protein may improve our understanding of the RVF infection cycle and may provide novel targets for antiviral therapies.

Materials and methods

Cells and virus

BSR-T7/5 cells were a generous gift of Dr. K. Conzelmann (Max-von Pettenkofer-Institut, Munchen, Germany). The cells were grown in Glasgow modified Eagle medium supplemented with 10% fetal calf serum, 1% nonessential amino acids, 5% tryptose phosphate broth and Geneticin (1 mg/mL). MVA-T7pol was a generous gift of Dr. G. Sutter (GSF, Germany).

Plasmids

RVF T7-7 was a generous gift of Dr. M. Parker at the U. S. Army Medical Research Institute of Infectious Diseases (Ft. Detrick, MD), and consists of nucleotides 20-3767 (message-sense) of the M segment for the ZH501 strain of RVF virus in pSP73. RVFM is a full-length clone of the M segment from the ZH501 strain of RVF virus, and was constructed in several steps. A unique *Xho*I site was created in RVF T7-7 utilizing QuikChange mutagenesis kit and oligos G1*Xho*I: ATCACCACTTGCTCTCTCGAGGGTGTTAACACC AND G1*Xho*Irc: GGTGTTAACACCCTCGAGAGAGCAAGTGGTGAT, thus creating RVF T7-7*Xho*I. RVFMSTART5: GTCGACTAATACGACTCACTATAGACACAAAGACGGTGCATTAAATGTATGTTTTATTAACAATTC and RVFMSTART3: ATGTGCAGTGCTGAGTTGGCCATC were used to amplify the 5′ end of the M segment utilizing RVF T7-7 as the template, the resulting fragment was cloned into pCR4 (pSRG252). RVFMTERM5: CAACAGTTGTGAATCCAAAATC and RVFMTERM3: ACACAAAGACCGGTGCAACTTC were utilized to amplify the 3′ end of the M segment utilizing ZH501 RNA as the template in a two step RT-PCR reaction the resulting fragment was cloned into pCR4 (RVFMterm). RIBOT7-5: TTGCACCGGTCTTTGTGTGGGTCGGCATGGCATCTCCAC and RIBOT7-3: CTGGAATTCGGCTTAAAAAACCCC were used to amplify the HDV ribozyme and T7 terminator and the resulting fragment was cloned into pCR4 (pSRG253). The *Sal*I/*Msc*I fragment from pSRG252 was subcloned into the *Sal*I/*Msc*I sites of RVF T7-7*Xho*I. The resulting plasmid was then cut with *Sal*I/PflMI and RVFMterm plasmid with PflMI/*Eco*RI and both fragments subcloned into pSP64. The resulting plasmid was then cut with *Age*I/*Eco*RI and the *Age*I/*Eco*RI fragment of pSRG253 was subcloned in, thus generating a full-length clone of the M segment, pSRG255. A FLAG epitope was inserted into Gc by digesting pSRG255 with *Xho*I then inserting the epitope sequence (*Sal*I ends) by ligation with oligos G1FLAGTS: TCGACTACAAAGACCATGACGGTGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATG and G1FLAGBS: TCGACATCGTCATCCTTGTAGTCGATGTCATGATCTTTATAATCACCACCGTCATGGTCTTTGTAG. The ΔNSm plasmid was made in an analogous fashion to pSRG255, except that RVFM-NSm5: GTCGACTAATACGACTCACTATAGACACAAAGACGGTGCATTAAATGGCAGGGATTGCAATGACAG was used, instead of RVFMSTART5 to amplify the 5′ end of the M segment, thus creating pSRG254. The final assemblage of ΔNSm plasmid was analogous to RVFM, and the final plasmid is designated pSRG257. The FLAG epitope was inserted into pSRG257 utilizing G1FLAGTS and G1FLAGBS, thus generating pSRG259. Putative signal peptidase cut sites were ablated in the RVFM and ΔNSm plasmids by mutagenizing the alanine residue in the −1 position to arginine with QuikChange mutagenesis kit and G2SPMUT: TTGGCACCTGTTGTTTTTAGAGAAGACCCCCATCTCAGA and G2SPMUTRC: TCTGAGATGGGGGTCTTCTCTAAAAACAACAGGTGCCAA for the Gn site and G1SPMUT: ATTGTCTCATATGCATCAAGATGTTCAGAACTGATTCAG and G1SPMUTRC: CTGAATCAGTTCTGAACATCTTGATGCATATGAGACAAT for the Gc site. The plasmid encoding Gc with a FLAG epitope was generated by amplification of RVFMGcFLAG with Gc5: GGATCCATGTATAGCACATACCTGATGTTATTA and Gc3: CATCTAAAGTATTATTGCATAAG, the resulting product was cloned into the *Bam*HI/*Eco*RI sites of pcDNA1.1. The plasmid encoding Gn has been described previously (Gerrard and Nichol, 2002).

Materials

A QuikChange mutagenesis kit was obtained from Stratagene (La Jolla, Calif.). MEM without cysteine and methionine was obtained from MP Biomedical (Burlingame, CA). TransIT-LT1 was obtained from Mirus Corporation (Madison, WI). L-cysteine, L-methionine and mouse anti-FLAG were obtained from Sigma (St. Louis, MO). ^{14}C radiolabeled molecular weight markers and protein G-sepharose were obtained from Amersham Pharmacia Biotech (Piscataway, NJ). ^{35}S-labeled cysteine was obtained from NEN Life Science Products (Boston, MA). Monoclonal antibodies recognizing Gn have been described previously (Gerrard and Nichol, 2002).

Signal peptidase cleavage site predictions

Prediction of signal peptides and signal peptidase cleavage positions were performed using the hidden Markov model and neural network model prediction programs at SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/). For the prediction of the signal peptide upstream of Gn, we used four amino acid sequence inputs. In all cases, amino acid numbering begins at the methionine at which translation initiates. The sequences utilized are as follows: (1) the region beginning at the second methionine within the NSm region (amino acid 1 in NSm2) and ending at amino acid 122 (Fig. 2C), (2) the sequence beginning at the fourth methionine and ending at amino acid 25 (Fig. 2D), (3) the sequence beginning at the fifth methionine and ending at amino acid 19 (data not shown) and, (4) the sequence encompassing the cytoplasmic tail of Gn, hydrophobic region upstream of Gc (data not shown). The decision to mutate the amino acid in the −1 position of the signal peptide cleavage site was based on the data presented in Nielsen et. al (Nielsen et al., 1997). There is a marked preference for small and uncharged amino acids (e.g. alanine, glycine, serine, cysteine) at the −1 position in the signal peptidase cleavage site (Nielsen et al., 1997), we therefore chose to mutate the alanine at this position to arginine in the predicted cleavage sites upstream of Gn and Gc.

Immune precipitations

BSR-T7/5 cells were grown in 6-well plates to 80% confluency then infected with MVA-T7pol at an MOI of 5. At 1 h post-infection, the inoculum was removed, the cells rinsed in PBS and 1 mL of fresh media was added, cells were then transfected with 1 μg/well of expression plasmid

utilizing TransIT-LT1. At 16 h post-infection, cells were starved in labeling media (MEM without cysteine, 2% dialyzed FCS) for 30 min, then were labeled for 5 min with 100 μCi ^{35}S-cyteine in labeling media. Following labeling, excess unlabeled cysteine was added and cells were harvested in Triton Lysis buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 5 mM EDTA, 1% Triton TX-100) at the indicated chase times. Extracts were then centrifuged at 14,000 g for 5 min, the supernatant was removed and SDS was added to a final concentration of 0.1%. Samples were pre-cleared by incubation with protein G-sepharose, followed by centrifugation at 14,000 g for 5 min in a microfuge. The supernatant was removed to a fresh microfuge tube containing 2 μL of the appropriate monoclonal antibody. Samples were incubated for 1 h on wet ice, at which point protein G-sepharose was added and the incubation was continued for an additional hour at 4 °C with end-over-end agitation. Immune complexes were harvested by centrifugation and washed three times with Triton Lysis buffer with 0.1% SDS. Following the final wash immune complexes were solubilized in SDS-PAGE sample buffer with 100 mM DTT. Immune precipitated complexes were resolved by SDS-PAGE (midi-gel), the gel was then dried and placed in a PhosphoImager cassette. Screens were then scanned with a Storm PhosphoImager and band intensities were quantified using ImageQuant software (Amersham Pharmacia Biotech, Piscataway N.J.).

Acknowledgments

We thank Dr. G. Sutter, Dr. K Conzelmann and Dr. M. Parker for their generous gifts of reagents. We thank Dr. T. Ksiazek and Dr. P. Rollin for their support and critical comments during the course of this study. We also thank Dr. B. Levi for critical reading of the manuscript.

References

Anonymous, 2000. Outbreak of Rift Valley fever—Yemen, August–October 2000. MMWR Morb. Mortal. Wkly. Rep. 49, 1065–1066.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795.

Chen, S.Y., Compans, R.W., 1991. Oligomerization, transport, and Golgi retention of Punta Toro virus glycoproteins. J. Virol. 65, 5902–5909.

Collett, M.S., 1986. Messenger RNA of the M segment RNA of Rift Valley fever virus. Virology 151, 151–156.

Collett, M.S., Purchio, A.F., Keegan, K., Frazier, S., Hays, W., Anderson, D.K., Parker, M.D., Schmaljohn, C., Schmidt, J., Dalrymple, J.M., 1985. Complete nucleotide sequence of the M RNA segment of Rift Valley fever virus. Virology 144, 228–245.

Easterday, B.C., 1965. Rift valley fever. Adv. Vet. Sci. 10, 65–127.

Gad, A.M., Hassan, M.M., el Said, S., Moussa, M.I., Wood, O.L., 1987. Rift Valley fever virus transmission by different Egyptian mosquito species. Trans. R. Soc. Trop. Med. Hyg. 81, 694–698.

Gargan II, T.P., Clark, G.G., Dohm, D.J., Turell, M.J., Bailey, C.L., 1988. Vector potential of selected North American mosquito species for Rift Valley fever virus. Am. J. Trop. Med. Hyg. 38, 440–446.

Gerrard, S.R., Nichol, S.T., 2002. Characterization of the Golgi retention motif of Rift Valley fever virus G(N) glycoprotein. J. Virol. 76, 12200–12210.

Godet, M., L'Haridon, R., Vautherot, J.F., Laude, H., 1992. TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. Virology 188, 666–675.

Gro, M.C., Di Bonito, P., Fortini, D., Mochi, S., Giorgi, C., 1997. Completion of molecular characterization of *Toscana phlebovirus* genome: nucleotide sequence, coding strategy of M genomic segment and its amino acid sequence comparison to other phleboviruses. Virus Res. 51, 81–91.

Holland, J., Domingo, E., 1998. Origin and evolution of viruses. Virus Genes 16, 13–21.

Jupp, P.G., Cornel, A.J., 1988. Vector competence tests with Rift Valley fever virus and five South African species of mosquito. J. Am. Mosq. Control Assoc. 4, 4–8.

Jupp, P.G., Kemp, A., Grobbelaar, A., Lema, P., Burt, F.J., Alahmed, A.M., Al Mujalli, D., Al Khamees, M., Swanepoel, R., 2002. The 2000 epidemic of Rift Valley fever in Saudi Arabia: mosquito vector studies. Med. Vet. Entomol. 16, 245–252.

Kakach, L.T., Wasmoen, T.L., Collett, M.S., 1988. Rift Valley fever virus M segment: use of recombinant vaccinia viruses to study Phlebovirus gene expression. J. Virol. 62, 826–833.

Kakach, L.T., Suzich, J.A., Collett, M.S., 1989. Rift Valley fever virus M segment: phlebovirus expression strategy and protein glycosylation. Virology 170, 505–510.

Keegan, K., Collett, M.S., 1986. Use of bacterial expression cloning to define the amino acid sequences of antigenic determinants on the G2 glycoprotein of Rift Valley fever virus. J. Virol. 58, 263–270.

Kozak, M., 2002. Pushing the limits of the scanning mechanism for initiation of translation. Gene 299, 1–34.

Meegan, J.M., Bailey, C.L., 1988–1989. Rift Valley fever. In: Monath, T.P. (Ed.), The Arboviruses: Epidemiology and Ecology, vol. IV. CRC Press, Boca Raton, FL, pp. 51–76 (Notes: LC Control Number: 87029992, Includes bibliographies and indexes).

Melin, L., Persson, R., Andersson, A., Bergstrom, A., Ronnholm, R., Pettersson, R.F., 1995. The membrane glycoprotein G1 of Uukuniemi virus contains a signal for localization to the Golgi complex. Virus Res. 36, 49–66.

Moya, A., Holmes, E.C., Gonzalez-Candelas, F., 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev., Microbiol. 2, 279–288.

Nichol, S.T., 2001. Bunyaviruses. In: Fields, B.N., Knipe, D.M., Howley, P.M., Griffin, D.E. (Eds.), Fields' Virology. Lippincott Williams and Wilkins, Philadelphia, pp. 1603–1633.

Nielsen, H., Krogh, A., 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. Proc. Int. Conf. Intell. Syst. Mol. Biol. 6, 122–130.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1–6.

Persson, R., Pettersson, R.F., 1991. Formation and intracellular transport of a heterodimeric viral spike protein complex. J. Cell Biol. 112, 257–266.

Pettersson, R.F., Andersson, A., Melin, L., 1995. Mapping a retention signal for Golgi localization of a viral spike protein complex. Cold Spring Harbor Symp. Quant. Biol. 60, 147–155.

Schmaljohn, C.S., Hooper, J.W., 2001. Bunyaviridae: the viruses and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M., Griffin, D.E. (Eds.), Fields' Virology. Lippincott Williams and Wilkins, Philadelphia, pp. 1581–1602.

Shi, X., Lappin, D.F., Elliott, R.M., 2004. Mapping the Golgi targeting and retention signal of Bunyamwera virus glycoproteins. J. Virol. 78, 10793–10802.

Shoemaker, T., Boulianne, C., Vincent, M.J., Pezzanite, L., Al-Qahtani, M.M., Al-Mazrou, Y., Khan, A.S., Rollin, P.E., Swanepoel, R., Ksiazek, T.G., Nichol, S.T., 2002. Genetic analysis of viruses associated with emergence of Rift Valley fever in Saudi Arabia and Yemen, 2000–01. Emerg. Infect Dis. 8, 1415–1420.

Spiropoulou, C.F., Goldsmith, C.S., Shoemaker, T.R., Peters, C.J., Compans, R.W., 2003. Sin Nombre virus glycoprotein trafficking. Virology 308, 48–63.

Suzich, J.A., Collett, M.S., 1988. Rift Valley fever virus M segment: cell-free transcription and translation of virus-complementary RNA. Virology 164, 478–486.

Suzich, J.A., Kakach, L.T., Collett, M.S., 1990. Expression strategy of a phlebovirus: biogenesis of proteins from the Rift Valley fever virus M segment. J. Virol. 64, 1549–1555.

Turell, M.J., Kay, B.H., 1998. Susceptibility of selected strains of Australian mosquitoes (Diptera: Culicidae) to Rift Valley fever virus. J. Med. Entomol. 35, 132–135.

Turell, M.J., Presley, S.M., Gad, A.M., Cope, S.E., Dohm, D.J., Morrill, J.C., Arthur, R.R., 1996. Vector competence of Egyptian mosquitoes for Rift Valley fever virus. Am. J. Trop. Med. Hyg. 54, 136–139.

Veijola, J., Pettersson, R.F., 1999. Transient association of calnexin and calreticulin with newly synthesized G1 and G2 glycoproteins of uukuniemi virus (family Bunyaviridae). J. Virol. 73, 6123–6127.

Wasmoen, T.L., Kakach, L.T., Collett, M.S., 1988. Rift Valley fever virus M segment: cellular localization of M segment-encoded proteins. Virology 166, 275–280.

Woods, C.W., Karpati, A.M., Grein, T., McCarthy, N., Gaturuku, P., Muchiri, E., Dunster, L., Henderson, A., Khan, A.S., Swanepoel, R., Bonmarin, I., Martin, L., Mann, P., Smoak, B.L., Ryan, M., Ksiazek, T.G., Arthur, R.R., Ndikuyeze, A., Agata, N.N., Peters, C.J., 2002. An outbreak of Rift Valley fever in Northeastern Kenya, 1997–98. Emerg. Infect Dis. 8, 138–144.