RIFT VALLEY FEVER VIRUS: MOLECULAR BIOLOGIC STUDIES OF THE M SEGMENT RNA FOR APPLICATION IN DISEASE PREVENTION

FINAL REPORT

MARC S. COLLETT, PhD

FEBRUARY, 1989

Supported by
U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21701-5012

Contract No. DAMD17-85-C-5226

Molecular Genetics, Inc.
10320 Bren Road East
Minnetonka, Minnesota 55343

DOD DISTRIBUTION
Approved for public release; distribution unlimited

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorize documents.

Reproduced From
Best Available Copy
The mission of this contract was to elucidate the molecular and biologic properties of the middle (M) genomic segment of Rift Valley fever virus (RVFV), and the apply the new-found basic research information toward strategies for RVF disease prevention. The molecular characteristics of the M segment messenger RNA were described, as were the gene products encoded therein. The single open reading frame within this RNA codes for four products; the two viral envelopes: glycoproteins G2 and G1, a glycosylated 78kd protein, and a nonglycosylated 14kd protein. The expression of the full complement of M segment-encoded proteins is complex and involves multiple translational initiation events giving rise to primary translation products which are cotranslationally processed to yield the mature proteins. Portions of the RVFV M segment were engineered into bacterial expression systems for the production of subunit protein immunogens and into recombinant vaccinia viruses for creation of live virus vaccine candidates. The bacterial products were marginally immunogenic, capable of protecting only a portion of immunized laboratory animals from RVFV challenge. Recombinant vaccinia viruses were 100% protective and represent viable vaccine candidates for RVF. The glycoprotein G2 was found to be the essential and sufficient protective immunogen. Important immunologic determinants on this protein were mapped. These investigations have provided a great deal of insight into the molecular and immunologic properties of the M segment and its products.
SUMMARY

The mission of this contract was to gain detailed and fundamental information regarding the molecular biologic and immunologic features of Rift Valley fever virus (RVFV). Specifically, our objectives involved elucidation of molecular and biologic properties of the middle (M) genomic segment of RVFV, and the application of new-found basic research information toward strategies for RVF disease prevention. The molecular characteristics of the M segment messenger RNA were described, as were the gene products encoded therein. The single open reading frame within this RNA codes for four products: the two viral envelope glycoproteins G2 and G1, a glycosylated 78kd protein, and a nonglycosylated 14kd protein. The expression of the full complement of M segment-encoded proteins is complex and involves multiple translational initiation events giving rise to primary translation products which are co-translationally processed to yield the mature proteins. Portions of the RVFV M segment were engineered into bacterial expression systems for the production of subunit protein immunogens and into recombinant vaccinia viruses for creation of live virus vaccine candidates. The bacterial products were marginally immunogenic, capable of protecting only a portion of immunized laboratory animals from RVFV challenge. Recombinant vaccinia viruses were 100% protective and represent viable vaccine candidates for RVF. The glycoprotein G2 was found to be the essential and sufficient protective immunogen. Important immunologic determinants on this protein were mapped. These investigations have provided a great deal of insight into the molecular and immunologic properties of the M segment and its products.
FOREWORD

In performing the recombinant DNA experiments described in this report, the investigators have abided by the National Institutes Of Health Guidelines for Research Involving Recombinant DNA Molecules (April 1982) and the Administrative Practices Supplements.

In conducting laboratory animal immunization experiments described in this report, the investigators adhered to the "Guide for the Care and Use of Laboratory Animals" of the Institute of Laboratory Animal Resources, National Research Council (DHEW Publication N.(NIH) 78-23, Revised 1978).
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report Documentation Page</td>
<td>2</td>
</tr>
<tr>
<td>Summary</td>
<td>2</td>
</tr>
<tr>
<td>Foreword</td>
<td>3</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>5</td>
</tr>
<tr>
<td>B. Molecular Biologic Features of the RVFV M Segment RNA</td>
<td>5</td>
</tr>
<tr>
<td>1. mRNA of the M Segment</td>
<td>5</td>
</tr>
<tr>
<td>2. Proteins encoded by the M Segment</td>
<td>6</td>
</tr>
<tr>
<td>3. Expression Strategy of the M Segment</td>
<td>7</td>
</tr>
<tr>
<td>C. Genetic Engineering of RVFV M Segment Sequences for Application in Disease Prevention</td>
<td>9</td>
</tr>
<tr>
<td>1. Bacterial Subunit Immunogens</td>
<td>9</td>
</tr>
<tr>
<td>2. Recombinant vaccinia viruses</td>
<td>9</td>
</tr>
<tr>
<td>3. Mapping Immunologic Determinants Important for Disease Prevention</td>
<td>10</td>
</tr>
<tr>
<td>D. Literature Cited</td>
<td>12</td>
</tr>
<tr>
<td>E. Publications Supported by this Contract</td>
<td>13</td>
</tr>
<tr>
<td>F. Personnel Receiving Contract Support</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1. Schematic representation of the RVFV M segment RNA and its protein products.</td>
<td>15</td>
</tr>
<tr>
<td>Distribution List</td>
<td>16</td>
</tr>
</tbody>
</table>
A. Introduction. This report summarizes work performed over the three year tenure of this contract. The mission of this contract was to gain detailed and fundamental information regarding the molecular biologic and immunologic features of Rift Valley fever virus (RVFV). Specifically, our original objectives involved elucidation of molecular and biologic properties of the middle (M) genomic segment of RVFV, and the application of new-found basic research information toward strategies for RVF disease prevention. The studies reviewed in this report show these objectives have indeed been met. We have gained a great deal of insight into the molecular and immunologic properties of the M segment and its gene products, and further have generated potentially useful candidate vaccines for RVF. And we had fun doing it! The collaborative atmosphere of the project with workers of the Virology Division of USAMRIID served not only to accelerate progress, but made the work enjoyable.

B. Molecular Biologic Features of the RVFV M Segment RNA. RVFV is a member of the Phlebovirus genus of the Bunyaviridae family. The genome of RVFV consists of three RNA segments designated L, M, and S. The S segment RNA encodes the viral nucleocapsid protein N and a nonstructural polypeptide NS, (1). Work carried out by us previously, in collaboration with the Virology Division of USAMRIID, resulted in the molecular cloning and complete nucleotide sequence determination of the M RNA segment of RVFV. This information, along with additional data, showed the M RNA encoded the two viral envelope glycoproteins G2 and G1, and possibly one or more yet unidentified polypeptides (2). Although no direct evidence is available, it is presumed the L segment RNA codes for the L protein, a component of the viral transcriptase.

1. mRNA of the M Segment. Members of Bunyaviridae appear to employ two fundamentally different strategies for gene expression: "negative sense" and "ambisense". From the nucleotide and protein sequence data (2), it was clear the M segment of RVFV employed the negative sense strategy: the single large open reading frame (ORF) encoding the viral glycoproteins was present in the virus-complementary polarity. To further understand the replication and expression of this genomic segment, its messenger RNA (mRNA) was characterized. Using S1 nuclease mapping and oligonucleotide primer extension procedures, we found the mRNA of the M segment was complementary to the virion RNA but lacked at its 3' end approximately 112 nucleotides present at the 5' end of the genomic RNA. The 5' end of the mRNA possessed all the sequences present at the 3' end of the M RNA, but was further extended beyond the end of the genome.
by about 12-14 nucleotides. These added nucleotides were heterogeneous in sequence (3). The lack of genomic sequences at the 3' end of the mRNA would prevent mRNA molecules from forming circular structures that genomic RNAs, which possess terminal complementary sequences, might form. This sequence difference between mRNAs and genomic (as well as antigenomic) RNAs may serve to distinguish molecules that participate in viral gene expression and those that are involved in the virus replication cycle. These sequence and/or structural differences may have important consequences in viral replicase recognition and transcription, as well as RNA encapsidation. The characteristics of the 5' end of the M segment mRNA are reminiscent of those of orthomyxoviruses (as well as some other viruses) that scavenge sequences from host cellular mRNAs for use in the initiation of viral mRNA transcription.

2. **Proteins encoded by the M Segment.** The large ORF of the M segment is capable of encoding 1197 amino acids, or about 133 kilodaltons (kd) of protein (2,4). This coding potential is in excess of that required for the two viral glycoproteins. After aligning the amino-terminal amino acid sequences of glycoproteins G2 and G1 within the ORF, the region from the first ATG codon of the ORF to the codon for the first amino acid of the mature glycoprotein G2 represented the extra coding potential. The likelihood of significant non-glycoprotein coding information between the end of the first glycoprotein (G2) and the beginning of the second (G1) was eliminated by the demonstration that antiserum generated against the amino acid sequence 16 to 27 residues preceding the mature glycoprotein G1 sequence reacted with glycoprotein G2 (5).

From the first ATG of the ORF to the beginning of the coding sequences for glycoprotein G2 there is a potential coding capacity of 17kd. Using sequence-specific antiserum reagents (anti-synthetic peptide antisera), we identified two additional gene products encoded by the RVFV M segment: a 78kd and a 14kd protein (5). The sequences represented in the 78kd protein began from the first methionine of the ORF and encompassed the entire preglycoprotein and glycoprotein G2 coding sequences. The sequence of the 14kd protein began from the second methionine residue of the ORF and represented only preglycoprotein sequences (Fig. 1). Analyses of these new proteins showed that although the 14kd protein possessed a potential site for N-linked glycosylation, it was not glycosylated (5). The 78kd protein possessed this same site, as well as the single N-linked site resident in the glycoprotein G2 sequence; both sites were occupied with glycan in the 78kd protein (6). Thus, the complete coding capacity of the RVFV M segment included the two viral envelope glycoproteins G2 and G1, a glycosylated 78kd protein, and a non-glycosylated 14kd protein. Immunofluorescent analyses of RVFV-infected cells revealed
3. Expression Strategy of the M Segment. To understand the expression strategy of the M segment, and the mechanisms involved in the biogenesis of its four gene products, we employed two genetically manipulable systems for the surrogate expression of RVFV sequences: cell-free transcription-translation for in vitro studies and recombinant vaccinia viruses for analyses in vivo. Our results indicate the expression strategy of the RVFV M segment is surprisingly complex.

The fact that the M segment mRNA consists of a single ORF suggested biogenesis of the mature proteins proceeds by proteolytic processing of a primary translation product. Indeed, the cell-free translation of M segment mRNA-like transcripts revealed a primary polyprotein product encompassing the entire ORF which was co-translationally processed in vitro in the presence of microsomes to yield mature proteins (d). However, on occasion, we were able to resolve what appeared to be multiple, closely-spaced primary translation products in this cell-free system. This led us to consider that multiple translational initiation events within the same ORF might also be involved in the biogenesis of M segment proteins; there are four additional in-phase ATG codons downstream of the first ATG within the preglycoprotein region of the ORF (Fig. 1). In fact, based on sequence-specific antibody recognition, the 14kd protein neatly lacked reactivity with antisera directed at sequences between the first and second ATG codons of the ORF, but did react with antibodies to the 12 amino acid sequence between the second and third ATGs; the 78kd protein reacted with immune reagents for both regions (5, 8). Thus, there were two possible mechanisms for the biogenesis of the 14kd protein: this polypeptide arose as a proteolytic processing product of the 78kd protein, or it represented the product of an independent translational initiation event at the second in-phase ATG codon of the ORF. To investigate these two alternatives, we exploited recombinant vaccinia virus technology. We had shown during the course of this contract work that when the complete RVFV M segment ORF was inserted into vaccinia virus under the control of a vaccinia virus promoter, M segment gene expression precisely mimicked that seen in authentic RVFV-infected cells. Cells infected with such a recombinant vaccinia virus correctly expressed, processed, and transported all M segment proteins (5, 7). Therefore, we were able to use this expression system to explore M segment protein biogenesis. Of the two alternatives mentioned above for the production of the 14kd protein, pulse-chase experiments using recombinant vaccinia virus infected cells
failed to suggest a precursor-product relationship between the 78kd and 14kd proteins (9). However, by combining site-directed mutagenesis techniques with the recombinant vaccinia virus system, we showed the first ATG codon was necessary for synthesis of the 78kd protein, but not for the 14kd protein, and the second ATG was essential for 14kd protein production, but not for 78kd protein synthesis (9,10).

These data show the M segment of RVFV employs two different in-phase translational initiation codons to produce two distinct polypeptide products, the 78kd and 14kd proteins. Cell-free translation studies indicated the primary product of M segment mRNA translation to be protein species the size expected of a polypeptide encompassing the entire ORF (8). Taken together, these data would predict there to be two primary translation products from the same M segment mRNA: one initiating from the first ATG codon and extending to the termination codon of the ORF (1197 amino acids) and a second beginning from the second in-phase ATG and continuing to the end of the ORF (1159 amino acids). Using sequence-specific antiserum reagents and gel analyses of sufficient resolution, we have recently been able to clearly resolve these two primary translation products (unpublished data). Thus, biogenesis of the complete complement of M segment-encoded proteins involves both multiple translational initiation events and co-translational proteolytic processing of precursor polyproteins.

Although the above information explains the mode of production of the 78kd and 14kd proteins, they do not address the biogenesis of the viral glycoproteins G2 and G1. Pulse-chase experiments showed no indication the 78kd protein served as a precursor to glycoprotein G2 (9). Moreover, elimination of either the first or the second ATG codon did not abrogate production of glycoproteins G2 and G1. Further elimination of the additional ATG codons within the preglycoprotein region by site-directed mutagenesis caused a progressive decrease in glycoprotein G2 expression but failed to significantly affect the production of glycoprotein G1 (10). We conclude the biogenesis of glycoprotein G2 is dependent on translation start sites within the preglycoprotein region, but does not involve use of the first ATG codon of the ORF, and the production of glycoprotein G1 appears to be largely independent of all ATG codons that precede the mature glycoprotein coding sequences. Unfortunately, our data to date do not completely elucidate the mechanisms used for the production of glycoproteins G2 and G1, but they do clearly indicate the biogenesis of the two envelope glycoproteins of RVFV proceeds along two distinctly different pathways.

The post-translational consequences of the two-site translational initiation strategy employed for the expression of the 78kd and 14kd proteins were investigated in an attempt to ascribe to it a functional role. We found translational initiation from the first ATG, and therefore incorporation into protein of the 37 amino acids preceding the second in-
phase ATG codon, appears to predetermine utilization of the N-linked glycosylation site within the preglycoprotein region of the ORF, and furthermore, precludes proteolytic cleavage at the preglycoprotein-glycoprotein G2 junction. On the other hand, translation from the second ATG codon results in the failure to use the preglycoprotein glycosylation site, but does allow proteolytic cleavage at the amino terminus of mature glycoprotein G2 to yield the 14kDa protein. However, proteolytic cleavage at this junction is independent of glycosylation at the preglycoprotein site (6).

Therefore, the amino acid sequence between the first and second ATG codons is implicated in exerting a profound influence on subsequent protein glycosylation and proteolytic processing. We speculate the presence or absence of these 37 amino acids may affect the conformation, or possibly intracellular transit, of the resultant polypeptide. This affect in turn influences what subsequent modifications may take place on the polypeptide. Thus, use of the two-site translational initiation expression strategy by this phlebovirus M segment serves as a mechanism for controlling post-translational protein modifications.

C. Genetic Engineering of RVFV M Segment Sequences for Application in Disease Prevention. Two approaches toward developing a subunit vaccine for RVF were pursued during the period of this contract: bacterial production of subunit immunogens and use of live recombinant vaccinia viruses.

1. Bacterial Subunit Immunogens. Bacterial expression plasmids were designed and constructed so that when introduced into E. coli they resulted in the production of RVFV glycoprotein analogue polypeptides. These proteins were partially purified from the bacterial cells and evaluated as subunit immunogens. Numerous expression constructs were made, formulated in various manners, and used to immunize laboratory mice. The sum of a large amount of data indicated mice developed only marginal titers of virus-neutralizing antibodies in response to these immunogens as measured by a plaque reduction neutralization test. However, with select polypeptide analogues, a significant proportion (50-70%) of immunized mice were protected from a lethal RVFV challenge (11). Protective immunity was observed only with glycoprotein G2 analogues, not glycoprotein G1 derivatives. Attempts to increase the protective immunogenicity of E. coli-produced materials were unsuccessful. Due to these difficulties, and to the increasing promise of our recombinant vaccinia virus work, we discontinued efforts in this area early on in the contract term.

2. Recombinant vaccinia viruses. Our second approach to RVFV vaccine development involved the construction, characterization, and evaluation as immunogens of live recombinant RVFV-vaccinia viruses. Initially, we constructed
a recombinant virus incorporating the coding region for both viral glycoproteins and showed that administration of this live virus to mice resulted in the development of high virus-neutralizing antibody titers, as well as almost complete protection (97.5%) against lethal RVFV challenge (11). Additional recombinant viruses were constructed, each expressing a slightly different portion of the M segment ORF. The sum of the results from mouse protection experiments showed glycoprotein G2 was the critical and sufficient protective immunogen: immunization with recombinant vaccinia viruses expressing only glycoprotein G2 was capable of protecting 100% of the animals. It was further established that virus neutralizing antibody activity did not correlate with protection of animals from disease. Sera from animals immunized with recombinant viruses expressing only glycoprotein G1 showed high virus neutralizing titers, but these mice did not survive a lethal RVFV challenge. It was clearly demonstrated that recombinant vaccinia viruses are viable live virus vaccine candidates for RVF.

3. Mapping Immunologic Determinants Important for Disease Prevention. The above animal protection data clearly indicated the importance of glycoprotein G2. We went on to further characterize important immunologic determinants on this polypeptide.

Monoclonal antibodies (mAbs) to glycoprotein G2 had been generated by workers of the Virology Division of USAMRIID. Several of these that had virus-neutralizing activity in vitro were also capable of protecting animals from lethal RVFV challenge upon their passive transfer to mice. Clearly, these mAbs recognized important determinants of the G2 glycoprotein. To identify the epitopes defined by three of these mAbs, we used a bacterial expression cloning system. The three antigenic determinants were ultimately localized along the glycoprotein G2 sequence to small distinct peptide regions of between 11 and 34 amino acids in length (12). Although the antigenic nature of these obviously important sequences was established, their immunogenicity, and therefore their potential utility for vaccine purposes, has not been.

Since we had shown glycoprotein G2 was the sufficient immunogen for elicitation of protective immunity, we wanted to further define and characterize the "protective" determinants of glycoprotein G2. We constructed a series of recombinant vaccinia viruses that possessed glycoprotein G2 gene sequences that were progressively truncated from the COOH-terminus. Each of these viruses expressed the expected truncated polypeptide in virus-infected cells. In work done by J.M. Dalrymple and S.E. Hasty of USAMRIID, outbred and inbred mice were immunized with each of the viruses in this series and then challenged with virulent RVFV. Immunization with the recombinant viruses expressing the complete G2 glycoprotein and a 75 amino acid truncated version were 100%
protective. However, the next virus in the series, having 186 amino acids of the COOH-end of glycoprotein G2 deleted, showed no ability to protect mice from lethal RVFV challenge (13). These results are provocative in view of previously mapped antigenic determinants reactive with protective monoclonal antibodies (12): all of the mapped domains were represented in the glycoprotein G2 sequences resident in the non-protective virus. Possible interpretations of these data include: (i) the presence of important protective epitopes near the carboxy terminus of G2, (ii) the requirement for a complete COOH-terminus of G2 for proper cellular processing of antigenic viral components, or (iii) the carboxy terminus of glycoprotein G2 is important for the presentation and/or conformation of epitopes located within the amino proximal portion of the molecule.
D. Literature Cited


E. Publications Supported by this Contract


F. Personnel Receiving Contract Support

Susan Belzer, BS
Marc S. Collett, PhD
Conrad Gold, BS
John Humphrey, BS
Laura Torborg Kakach, BS
Kathy Keegan, BS
Richard Krzyzek, PhD
Dawn Newman, BS
Gary Opperman, BS
P. Sridhar, PhD
JoAnn A. Suzich, PhD
Terri L. Wasmoen, PhD
Figure 1. Schematic representation of the RVFV M segment RNA and its protein products. The RNA is presented in the viral-complementary polarity. Nucleotide coordinates for the beginning of the mature G2 (481) and G1 (2092) glycoproteins, the termination codon for the ORF (3612), and the end of the genome (3885) are given. The vertical tick marks within the preglycoprotein region (PRE) represent the five in-phase ATG codons preceding the glycoprotein coding sequences. The lollipop figures depict the positions of the N-linked glycosylation recognition sequence: Asn-X-Thr/Ser. The four gene products and the regions within the ORF of their coding sequences are shown below the M segment.
<table>
<thead>
<tr>
<th>Copies</th>
<th>Recipient</th>
<th>Address</th>
</tr>
</thead>
</table>
| 5      | Commander | US Army Medical Research Institute of Infectious Diseases  
         |           | ATTN: SGRD-UlZ-M  
         |           | Fort Detrick, Frederick, MD 21701-5011 |
| 1      | Commander | US Army Medical Research and Development Command  
         |           | ATTN: SGRD-RMI-S  
         |           | Fort Detrick, Frederick, MD 21701-5012 |
| 2      | Defense Technical Information Center (DTIC)  
         |           | ATTN: DTIC-DDAC  
         |           | Cameron Station, Alexandria, VA 22304-6145 |
| 1      | Dean      | School of Medicine  
         |           | Uniformed Services University of the Health Sciences  
         |           | 4301 Jones Bridge Road  
         |           | Bethesda, MD 20814-4799 |
| 1      | Commandant | Academy of Health Sciences, US Army  
         |           | ATTN: AHS-CDM  
         |           | Fort Sam Houston, TX 78234-6100 |