Essential Amino Acids of the Hantaan Virus N Protein in Its Interaction with RNA

William Severson,1 Xiaolin Xu,2 Michaela Kuhn,3 Nina Senutovich,4 Mercy Thokala,5 François Ferron,6 Sonia Longhi,6 Bruno Canard,4 and Colleen B. Jonsson7*

Department of Entomology, Plant Pathology and Weed Science, P.O. Box 30003, MSC 3BE, New Mexico State University, Las Cruces, New Mexico 88003; Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, Alberta, Canada; Diplomand, Institute of Microbiology and Virology, University of Witten/Herdeck, Stockumstr. 10, 58448 Witten, Germany; NIH RISE Undergraduate Program, Department of Biology, New Mexico State University, Las Cruces, New Mexico 88003; Graduate Program in Biochemistry, Department of Entomology, Plant Pathology and Weed Science, P.O. Box 30003, MSC 3BE, New Mexico State University, Las Cruces, New Mexico 88003; Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Université Aix-Marseille I et II, ESIL, Campus de Luminy, 163 Avenue de Luminy, case 925, 13288 Marseille Cedex 09, France; and Department of Biochemistry and Molecular Biology, 2000 9th Ave. S, Southern Research Institute, Birmingham, Alabama 35205

Received 22 July 2004/Accepted 10 March 2005

The nucleocapsid (N) protein of hantavirus encapsidates viral genomic and antigenomic RNAs. Previously, deletion mapping identified a central, conserved region (amino acids 175 to 217) within the Hantaan virus (HTNV) N protein that interacts with a high affinity with these viral RNAs (vRNAs). To further define the boundaries of the RNA binding domain (RBD), several peptides were synthesized and examined for the ability to bind full-length S-segment vRNA. Peptide 195-217 retained 94% of the vRNA bound by the HTNV N protein, while peptides 175-186 and 205-217 bound only 1% of the vRNA. To further explore which residues were essential for binding vRNA, we performed a comprehensive mutational analysis of the amino acids in the RBD. Single and double Ala substitutions were constructed for 18 amino acids from amino acids 175 to 217 in the full-length N protein. In addition, Ala substitutions were made for the three R residues in peptide 185-217. An analysis of protein-RNA interactions by electrophoretic mobility shift assays implicated E192, Y206, and S217 as important for binding. Chemical modification experiments showed that lysine residues, but not arginine or cysteine residues, contribute to RNA binding, which agreed with bioinformatic predictions. Overall, these data implicate lysine residues dispersed from amino acids 175 to 429 of the protein and three amino acids located in the RBD as essential for RNA binding.

Hantaviruses, classified as emerging viruses, can cause two diseases when transmitted to humans, namely, hemorrhagic fever with renal syndrome and hantavirus pulmonary syndrome (5, 26). The prototype virus for hemorrhagic fever with renal syndrome is Hantaan virus (HTNV) and that for hantavirus pulmonary syndrome is Sin Nombre virus. These viruses are often referred to as Old World and New World viruses, respectively, because of their geographical distribution, which is limited by the ecological habitats of their rodent reservoirs (30). Members of Hantavirus, a genus within the Bunyaviridae family, have negative-strand, tripartite genomes that encode an RNA-dependent RNA polymerase (RdRp) (L segment), the nucleocapsid (N) protein (S segment), and the G1 and G2 glycoproteins (M segment) (31). G1 and G2, which are posttranslationally processed through the endoplasmic reticulum and the Golgi, are presented on the external face of the virion and facilitate the entry of the virus into the host cell. The three genomic viral RNAs (vRNAs) from a complex with N and possibly RdRp in the virion. These ribonucleocapsids, not named viral RNA, serve as templates for transcription and replication by RdRp in the cytoplasm of an infected host cell.

For many RNA viruses, assembly initiates with the binding of the N protein or a core protein to a unique encapsidation signal within the viral genome. This interaction promotes the oligomerization of the nucleocapsid protein, interactions with other viral proteins, and the subsequent formation of the virus particle. Over a decade ago, it was suggested that sequences or structures present in the 5′ end of the hantaviral genome could provide a point of nucleation for encapsidation by the N protein (27). Shortly thereafter, studies of Bunyamwera virus (BUNV), a member of the Orthobunyavirus genus within the family Bunyaviridae, suggested that the N protein encapsidates vRNA and cRNA (antigenomic RNA) but not mRNA or other nonviral RNAs (12). Using in vitro binding assays, we have shown previously that the HTNV N protein has a strong preference for vRNA compared to viral cRNA, mRNAs, and non-specific RNAs (32). A similar result was reported for the BUNV N protein (24). Both HTNV and BUNV N proteins have similar levels of affinity for their S-segment vRNA substrates, with dissociation constants of 53 nM and 80 nM, respectively (32). Specific cis-acting sequences were identified for both virus N proteins. The cis-acting elements that show the highest binding affinities for the HTNV and BUNV N proteins map to the 5′ ends of the vRNAs (24, 32). Furthermore, we

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, 2000 9th Avenue South, Southern Research Institute, Birmingham, AL 35205. Phone: (205) 581-2681. Fax: (205) 581-2093. E-mail: jonsson@sri.org.
have identified an RNA binding domain (RBD) within the HTNV N protein that maps to a central, conserved region (amino acids 175 to 217) (37). In addition to its interactions with viral RNA, the N protein has been shown to form trimers in vitro (16, 17) and in virions (1). It has been proposed that N protein trimers may assemble onto the viral RNA, followed by protein–protein interactions that promote the encapsidation of the entire RNAs or cRNAs (16). Recent experiments also suggested the involvement of a host cellular protein, the small ubiquitin-like modifier 1-conjugating enzyme 9 (Ubc9), in assembly (19). Additional macromolecular interactions are likely required for the assembly of the three nucleocapsids into the virion and the budding of this complex into the Golgi.

Our previous studies have outlined the location of an RNA binding domain (37) and a specific cis-acting element involved in HTNV nucleocapsid encapsidation and assembly (32). To map the boundaries and essential amino acids involved in the interaction of the HTNV N protein with RNA, single alanine substitutions were made within amino acids 175 to 217 of the full-length protein and in peptide mimics of the RBD. In addition, several different chemicals were used to modify specific functional groups of amino acids in the full-length HTNV N protein to identify amino acids that contribute to RNA binding. The results delineate the presence of a major RNA binding domain, with additional interactions being mediated by lysine residues scattered in the terminal half of the N protein.

**MATERIALS AND METHODS**

**Site-directed mutagenesis.** A Quick-Change site-directed mutagenesis kit (Stratagene, La Jolla, CA) was used to introduce single and double Ala substitutions into the full-length HTNV N protein, which is cloned into the pET2b bacterial expression vector (32). Single mutations were directly introduced into this plasmid with two synthetic oligonucleotide primers (available upon request). Each primer was designed to be complementary to the opposite strand of this plasmid with two synthetic oligonucleotide primers (available upon request).

**Protein purification.** The full-length HTNV N protein and N mutant proteins were purified as described previously (13, 37). Briefly, the wild-type HTNV N protein and mutant proteins were transformed into BL21(DE3) cells. Cells were grown overnight in 100 ml of LB medium containing 200 µg/ml ampicillin. After 12 to 14 h, the cells were centrifuged for 5 min at 5,000 rpm in a Sorval GS3 rotor. The pellet was resuspended in 40 ml of LB medium containing 200 µg/ml ampicillin and then inoculated into 1 liter of LB medium containing 200 µg/ml ampicillin. Cultures were grown for 2 h at 30°C before IPTG (isopropyl-β-D-thiogalactopyranoside) was added to initiate the final concentration of 1.2 mM to induce protein expression. After an induction time of 2 to 4 h, the cells were harvested (centrifugation for 5 min at 5,000 rpm in a Sorval GS3 rotor) and resuspended in 40 ml of extraction buffer at pH 8.0 (EB) (0.1 M sodium phosphate, 0.5 M sodium chloride, 0.1% Tween 20, 6 M guanidine-HCl). The resulting suspension was sonicated twice on ice using a Branson Sonifier until the solution was clear and not viscous. The extraction was continued for 2 h with shaking at room temperature. Soluble and insoluble materials were separated by centrifugation at 30,000 × g for 30 min. The supernatant was applied to a 5-ml column containing 2 ml of nickel-nitrilotriacetaate agarose resin (QiAGEN, Valencia, CA), which had been pre-equilibrated with 10 column volumes of EB, pH 8.0. The column was washed with 10 column volumes of buffer A (0.1 M sodium phosphate, 0.5 M sodium chloride, 1 M urea, pH 8.0, pH 6.5, and pH 5.0) and 1 M sodium chloride, 0.5 M urea). The N protein was eluted from the column using 10 column volumes of buffer A, pH 4.5. Three 1-ml fractions of each mutant protein and five 1-ml fractions of the wild-type protein were collected. Twenty microliters of each fraction was analyzed for protein content by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The fractions containing the desired protein were dialyzed in 5 days in buffer C (0.02 M HEPES, pH 7.5, 0.5 M sodium chloride, 10% glycerol) with the urea concentration decreased daily by half.

**Oligoribonucleotides and peptides.** The oligoribonucleotide vRNA 1-39 (5′-UAGUAGAUAGCUCCUAAAAGACAAUCAAGGAGGAAUC-3′) was synthesized on a 1 µM scale and purified by high-performance liquid chromatography by Integrated DNA Technologies, Inc. (Coraville, IA). The synthetic RNA was labeled at the 5′ terminus with [α-32P]UTP by the use of T4 polynucleotide kinase (New England Biolabs, Boston, MA) and was purified on Quick Spin columns (Roche, Indianapolis, IN). The peptides were synthesized by Sigma Genosys, who conducted all characterizations of the peptides. Peptides were analyzed by mass spectrometry and then examined for purity by liquid chromatography and high-performance liquid chromatography. Peptides were purified to 95% homogeneity.

**In vitro transcription of viral and nonviral RNAs.** Radiolabeled HTNV S-segment RNA transcripts were produced from the linear HTNV S/pGEM1 genome by using a MaxScript Sp RNA transcription kit (Ambion, Austin, TX) as described previously (32). pGEM7Zf was used as reverse transcription of a nonviral 67-nucleotide RNA used as a control RNA. An RNase kit (QIAGEN) was used to purify transcripts.

**Filter binding assay.** Binding reactions were done as previously described (32, 33). Briefly, purified HTNV N proteins or N deletions were serially diluted (22.7, 5.75, 2.52, 0.84, and 0.28 µM) in a final volume of 20 µl of binding buffer (40 mM HEPES, pH 7.4, 100 mM NaCl, 5% glycerol). Samples were incubated at 37°C for 5 min. 1 ng of [α-32P]UTP-labeled RNA was added to each reaction mix, and the reaction mixtures were incubated for an additional 10 min at 37°C. Signals were quantitated with a Storm Molecular Dynamics PhosphorImager and analyzed with ImageQuant, version 4.2, software.

**Gel electrophoresis mobility shift assay (GEMSA).** One nanogram of [32P]-radiolabeled vRNA S segment, prepared as described above, was incubated with 56 µM of purified N protein in binding buffer (40 mM HEPES, pH 7.4, 100 mM NaCl, 2.5 mM MgCl2, 5% glycerol) in a final reaction volume of 20 µl at 37°C for 10 min. Fifty units of RNase T1 (Ambion, Austin, TX) was added, and the reaction mixtures were incubated for an additional 10 min at 37°C. A microgram of heparin was then added, and the mixtures were incubated for an additional 15 min at 37°C. One microliter of sample buffer (30% glycerol and 0.2% bromophenol blue) was added to each reaction, and the reaction mixtures were loaded into a 6% acrylamide gel, separated by electrophoresis in 0.5% Tris-borate-EDTA buffer at 200 V (constant voltage) for 3.0 h, and visualized by autoradiography.

**Chemical modification of HTNV N protein.** To neutralize positive charges on lysine residues or block lysine residues, sulfo succinimidyl acetate (NHS) ( Pierce, Rockford, IL) was used. Chemical modification of Arg side chains was accomplished with 4-hydroxyphenacyl-2-glucoside (HPG) (Pierce). N-Succinimidyl iodoacetate (Pierce) was used as a cross-linking reagent to modify amine and sulfhydryl groups in close proximity. Lastly, N-ethylmaleimide (NEM) (Pierce) in combination with diithiothreitol (DTT) was used to modify Cys residues as described previously (14). To modify the appropriate residues, the chemical was added to a final concentration of 10 mM to aliquots of full-length HTNV N protein and incubated on ice for 30 min. To modify Cys residues with NEM, aliquots of HTNV N protein were treated with 10 mM NEM for 30 min on ice. After the 30-min incubation, DTT was added to a final concentration of 50 mM. Likewise, NEM was added to a final concentration of 10 mM to HTNV N protein treated with 50 mM DTT. Preincubated HTNV N proteins (56 µM) were then added to reaction mixtures as described above, and the complexes were analyzed by GEMSA.

**Sequence retrieval and HCA.** The sequences used for this study were obtained by using BLASTP (3) against the VAZYMoLo (9) and Trembl (available on request) databases. The N sequence accession numbers used for this study are as follows: VAZY697, Seoul virus (strain SR-11) (Sapporo rat virus); VQY838 (Tremb), Tchioupioula hantavirus; VAZy685, Dobrava virus; VAZy689, Han- taan virus; Q9WM3 (Trebl), Topografov hantavirus; VAZy135, Andes virus; VAZy683, Puumala virus; VAZy685, Dobrava virus; VAZy689, Han- taan virus; Q9WM3 (Trebl), Topografov hantavirus; VAZy135, Andes virus; VAZy683, Puumala virus; VAZy685, Dobrava virus; VAZy689, Han- taan virus; Q9WM3 (Trebl), Topografov hantavirus; VAZy135, Andes virus; VAZy683, Puumala virus; VAZy685, Dobrava virus; VAZy689, Han- taan virus; Q9WM3 (Trebl), Topografov hantavirus; VAZy135, Andes virus; VAZy683, Puumala virus; VAZy685, Dobrava virus; VAZy689, Han- taaan virus; Q9WM3 (Trebl), Topografov hantavirus; VAZy135, Andes virus; VAZy683, Puumala virus; VAZy685, Dobrava virus; VAZy689, Han- 

Downloaded from http://jvi.asm.org/ on September 8, 2017 by guest
bers of positively and negatively charged residues divided by the total number of amino acid residues. It was calculated using the program ProtParam at the EXPASY server (http://www.expasy.ch/tools). The mean hydrophobicity (H) is the sum of normalized hydrophobicities of individual residues divided by the total number of amino acid residues minus four residues (to take into account fringe effects in the calculation of hydrophobicity). Individual hydrophobicities were determined using the ProtParam program at the EXPASY server (http://www.expasy.ch/tools), using the options Hphi/Kyte and Doolittle and a window size of 5 and normalizing the scale from 0 to 1. The values computed for individual residues were then exported to a spreadsheet, summed, and divided by the total number of residues minus four to yield H.

For a given protein, R was then plotted against H. The charge/hydrophobicity diagram is divided into two regions by a line, which corresponds to the equation H = (R + 1.151)/2.785. If it falls in the left part of the diagram [where H < (R + 1.151)/2.785], a protein is predicted to be disordered, whereas it is predicted to be ordered if it is found on the right side of the diagram. The net charge/hydrophobicity method is only applicable to a protein or protein region if the region cannot be subdivided into shorter, structurally independent modules; otherwise, it might give conflicting results. The method was only validated for regions of >50 amino acids (aa) (35). An estimation of the error rate of the method can be drawn from previous work (34).

Prediction of unstructured regions. Sequences of proteins were submitted to the DISEMBL (18a) and PONDOR (18, 29) servers using the default parameters. Access to PONDOR was provided by Molecular Kinetics.

Amino acid composition analysis. The average sequence composition of globular proteins was taken from the “globular 3D” data set. If the average composition of an amino acid “X” in globular proteins is CGX and if CPX is the composition in X of a protein P, then deviation from the composition in X of globular proteins is defined for P as (CPX - CGX)CGX. An estimation of the error rate of the method can be drawn from previous work (34).

**RESULTS**

Bioinformatic analysis of the hantavirus N protein. In our previous work, HTNV N protein deletions were prepared in the N-terminal, C-terminal, or both regions of the N protein (37) and used to map a minimal RBD between amino acid residues 175 and 217. Intrinsically disordered regions have been shown to be involved in protein-RNA and protein-protein interactions (7, 36). To determine whether this was a characteristic of the 175-220 region of the hantavirus N protein, we carried out a bioinformatic analysis. We used BLASTP to retrieve 19 hantavirus N sequences homologous to that of HTNV N, with an overall identity of 48% and an overall similarity of 78%. A multiple sequence alignment including the most closely related sequences was then built. An examination of the HTNV N sequence using HCA pointed out the presence of three global domains (aa 95 to 141, 156 to 279, and 295 to 429) separated by two linkers (aa 142 to 155 and aa 280 to 294) and by a large disordered N-terminal domain (aa 1 to 94). This last domain is rich in lysine residues and displays a propensity for α-helix-induced folding (data not shown). HCA carried out on all the hantavirus N sequences shown in Fig. 1 revealed that they possess the same overall modular organization as HTNV N (data not shown). The 175-220 region displays a very typical and conserved pattern of hydrophobic clusters (Fig. 1) that suggests that this region is structured. However, the rather high content of charged residues, together with the relatively low content of hydrophobic residues compared to other globular regions, suggests that this region might be partly exposed to the solvent and thereby interact with a ligand/partner. Moreover, the relative enrichment of this region in lysine residues is reminiscent of RNA binding regions of RNA-dependent RNA polymerases and thus suggests that the putative ligand could be RNA.

In agreement with the results provide by HCA, a DISEMBL analysis of the HTNV, Andes virus (ANDV), and Bermejo virus N proteins shows that the 175-220 region is structured (data not shown). On the other hand, PONDOR gives a borderline disorder prediction for the 175-220 region of all Hantavirus N sequences shown in Fig. 1 (data not shown). The method using the hydrophobicity/net charge ratio gives a borderline prediction of native unfoldedness for the 170-220 region of HTNV N, while the corresponding region of ANDV N is predicted to be globular (data not shown). An analysis of the deviation in amino acid composition of the 175-220 regions of the HTNV and ANDV N proteins indicates that they are both depleted in order-promoting residues and enriched in Q, which is considered a disorder-promoting residue (data not shown). Therefore, this region is not consistently predicted to be disordered. Rather, different predictors give borderline results, suggesting that this region is structured and displays a high extent of flexibility.

Finally, PSI-PRED and Predict Protein both predict an α-helix encompassing residues 200 to 220 of HTNV N (but with a reliability of <50%) (Fig. 1). This prediction is in agreement with the HCA plot of HTNV N, which shows a potential α-helix in this region (Fig. 1).

In conclusion, the sequence properties of the region spanning residues 175 to 220 of HTNV N converge to suggest that it is a structured region with a high extent of flexibility, possibly allowing an interaction with a partner/ligand. The presence of numerous Lys residues suggests that the ligand may be RNA. For the following data, we employed three different experimental avenues to define the key amino acids in the HTNV N protein involved in RNA binding.

Reconstitution of the RNA binding activity of the Hantaan virus N protein in a peptide. To further confirm the function of this region in RNA binding, we designed and purchased synthetic peptides of the RBD (Table 1; Fig. 1). The largest peptide that Sigma-Genosys was able to synthesize was from amino acids 185 to 217 (Table 1). A filter-binding analysis of peptide 185-217 with a 5′-end vRNA (1-39) substrate showed it to have a $K_d$ of ~70 nM (data not shown). This value is similar to the $K_d$ published for the full-length HTNV N protein with S-segment viral RNA substrates (53 ± 8 nM) (32). Therefore, three arginine-to-alanine substitutions (R197A, R199A, and R213A) were made in this peptide (Table 2). Arg residues within the RBD were targeted for mutational studies, given their importance to a number of proteins in their RNA binding interactions.

The RNA binding ability of each peptide was measured by filter binding (Table 1). The percentage of full-length HTNV vRNA that was bound by each peptide is given compared to the amount bound by the full-length protein (Tables 1 and 2). Peptide 195-217 showed the greatest binding ability, with 94% of the RNA bound to the peptide relative to the N protein. Peptides 185-217 and 175-206 showed an approximately twofold reduction in binding, while approximately 1% of the vRNA was bound by peptides 205-217 and 175-186. These
results suggested that amino acids 195 to 217 contain essential amino acids for RNA binding (Table 1). While peptides 185-217 (R197A) and 185-217 (R199A) showed a 5-fold reduction in binding compared to the wild-type HTNV N protein (Tables 1 and 2), they only showed a 2.5-fold reduction compared to the 185-217 peptide. Hence, none of the R-to-A substitutions, individually, had a significant effect on RNA binding in the filter-binding assays.

**Chemical modification of HTNV N protein.** To further examine the contribution of specific classes of amino acids in the full-length HTNV N protein as they relate to HTNV vRNA binding, we used classical methods to modify residues within

<table>
<thead>
<tr>
<th>N peptide</th>
<th>Sequence</th>
<th>% vRNA bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>185-217</td>
<td>QSSMKAEEIT PGRYRTAVCG LYPAQIKARQ MIS</td>
<td>50 ± 5</td>
</tr>
<tr>
<td>195-217</td>
<td>PGRYRTAVCG LYPAQIKARQ MIS</td>
<td>94 ± 3</td>
</tr>
<tr>
<td>205-217</td>
<td>LYPAQIKARQ MIS</td>
<td>1 ± 0.1</td>
</tr>
<tr>
<td>175-206</td>
<td>KHLYVSLPNA QSSMKAEEIT PGRYRTAVCG LYP</td>
<td>44 ± 2</td>
</tr>
<tr>
<td>175-186</td>
<td>KHLYVSLPNA QSS</td>
<td>0.9 ± 0</td>
</tr>
</tbody>
</table>

* A 56 μM concentration of each peptide was used for the analysis.

* Percentages of vRNA bound to N peptides are from three separate experiments performed in duplicate (± SD) compared to the HTNV N protein, set to 100%.
the full-length protein. The RNA binding activity of the HTNV N protein after chemical modification of Cys, Lys, or Arg or exposure to a cross-linking agent was assayed by GEMSA with the full-length HTNV S-segment RNA.

To investigate the effect of modifications of Cys residues on RNA binding, we treated the HTNV N protein with NEM. The HTNV N protein contains five Cys residues, one of which is in the RBD (C203) according to methods used previously in the laboratory to chemically probe the function of Cys (14). In the HTNV N protein, the other four Cys residues are located at aa positions 244, 293, 315, and 319. We noted no inhibition of RNA binding by the HTNV N protein when it was treated with NEM (Fig. 2, lanes 3 and 4). The levels of RNA-N protein complexes were similar for NEM-treated and untreated HTNV N proteins. The overexposed band at the bottom of Fig. 2 represents the free probe. A second band comigrating with the second band in the no-protein control lane is RNA secondary structure, and we have repeatedly noted this with no-protein controls. We have tried heating the RNA probe to 75°C for 15 min to eliminate possible secondary structure formation, but we always observe this band with the no-protein control. The predicted structure of the viral S-segment RNA has large areas of sequence that are double-stranded, and we suspect this region renatures to its secondary structure under GEMSA conditions (15). For the initial cloning of hantavirus genes, strong denaturants such as methylmercury were added to facilitate cloning (Connie Schmaljohn, personal communication), which suggests that the viral RNA has an inherent propensity for strong secondary structures. The third band is the N protein-bound RNA. We suggest that the slowest-migrating band may indicate additional complexes such as the trimeric form of the protein binding to the RNA (16, 17, 21, 22). Similar GEMSA patterns were observed throughout our studies and are shown in Fig. 3 to 5.

As noted in the bioinformatic analysis, the RBD is rich in Lys residues. Eleven percent of the aa residues in the full-length HTNV N protein are either Lys or Arg. Within the HTNV N protein RBD (aa 175 to 217), 6 of the 43 aa (14%) are Lys or Arg residues. When we treated the HTNV N protein with NHS, which neutralizes the positive charge on Lys residues, there was a complete inhibition of RNA binding (Fig. 2, lanes 5 and 6). However, the treatment of HTNV N protein

![FIG. 2. GEMSA of chemical modification of amino acid residues in HTNV N. Binding affinities were examined by GEMSA, in duplicate, with in vitro-transcribed full-length HTNV S-segment RNA and the expressed wild-type N protein (lane 2). Samples were treated with NEM (lanes 3 and 4), NHS (lanes 5 and 6), HPG (lanes 7 and 8), and N-succinimidyl iodoacetate (lanes 9 and 10). A no-protein control is presented in lane 1. A 56 μM concentration of HTNV N protein was treated with a 10 mM final concentration of each chemical modification reagent and then incubated with the radiolabeled vRNA S segment. The protein-vRNA complexes were separated from free RNA by 6% nondenaturing polyacrylamide gel electrophoresis (PAGE) as described in Materials and Methods.](http://jvi.asm.org/)

![FIG. 3. GEMSA of chemical modification of wt and truncated forms of HTNV N. Binding affinities were examined by GEMSA with in vitro-transcribed full-length HTNV S-segment RNA and the expressed wt N protein (lanes 2 and 9), ΔN209 (lanes 3 and 10), NP175-217 (lanes 4 and 11), NP175-270 (lanes 5 and 12), NP175-300 (lanes 6 and 13), and ΔΔ255 (lanes 7 and 14). A no-protein control is presented in lanes 1 and 8. A 56 μM concentration of HTNV N protein or a deletion (lanes 9 to 14) was treated with a 10 mM final concentration of NHS and incubated with the radiolabeled vRNA S segment. The reaction mixtures were loaded onto a 6% nondenaturing PAGE gel, and the protein-vRNA complexes were separated from free RNA by gel electrophoresis as described in Materials and Methods.](http://jvi.asm.org/)
with HPG, which reacts with the guanidyl group of Arg, had no effect on RNA binding (Fig. 2, lanes 7 and 8). Similarly, chemical modification with succinimidyl iodoacetate (Fig. 2, lanes 9 and 10), which cross-links proteins via an amine to thio linkage, showed no reduction in band shift intensity.

To further map the region where critical Lys residues reside, five previously prepared truncated constructs that removed conserved regions in the N-terminal, C-terminal, or both regions of the N protein (37) were subjected to chemical modification by NHS and examined by GEMSA analysis (Fig. 3). NΔ209 is devoid of the first 209 N-terminal amino acids, while CΔ255 is devoid of the last 255 C-terminal amino acids. NP175-217, NP175-270, and NP175-300 have both N-terminal and C-terminal deletions. In the presence of the chemical modifier NHS, there was a complete inhibition of RNA binding of the HTNV N protein, NΔ209 (Fig. 3, lane 10), NP175-217 (lane 11), NP175-270 (lane 12), and NP175-300 (lane 13) compared to untreated proteins (Fig. 3, lanes 3 to 6). However, there was only a 6.5-fold reduction in band shift intensity with CΔ255 (Fig. 3, compare lane 14 to lane 7), suggesting that Lys residues located between aa 175 and 429 contribute to RNA binding.

What are the critical amino acids in the Hantaan virus N protein that are used for its interaction with vRNA and cRNA?

To map the critical amino acids within the RBD, we created single amino acid substitutions in the full-length N protein from aa 175 to 217. We targeted amino acids that are known to interact with nucleic acids, such as K, R, T, S, D, and E. The 31-amino-acid region encompassing residues 175 to 217 does not have any W or F residues, but it does have several Y residues, which can interact through the aromatic ring via stacking interactions. Hence, these were also targeted for mutational analysis. Mutant N proteins were purified and examined for their ability to bind viral and nonviral RNAs compared to that of the wt N protein.

There was a reduction in binding of mutants with the single amino acid substitutions E192A, Y206A, and S217A compared to wt N protein (Fig. 4). The percentage of HTNV vRNA bound to mutant N proteins E192A and S217 was 16%, a 6.2-fold decrease compared to wt N protein (Fig. 4, lanes 10 and 20; Table 3). The Y206A mutant showed a 4.5-fold reduction in signal intensity (Fig. 4, lane 16; Table 3). The remaining 15 mutant N proteins had approximately the same signal intensities as the wt N protein (Fig. 4; Table 4). Three amino acids, namely, a glutamic acid (E192), a tyrosine (Y206), and a serine (S217) in the RBD, were identified as being essential for RNA recognition.

While single mutations in E192, Y206, and S217 reduced RNA binding, none of the mutants were devoid of binding.

Table 3. GEMSA analysis of amino acid substitutions in the HTNV N protein RBD

<table>
<thead>
<tr>
<th>Amino acid substitution(s)</th>
<th>% vRNA bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTNV N (WT)</td>
<td>100</td>
</tr>
<tr>
<td>Y178A</td>
<td>100</td>
</tr>
<tr>
<td>E192A</td>
<td>16</td>
</tr>
<tr>
<td>Y206A</td>
<td>22</td>
</tr>
<tr>
<td>Y178A/Y206A</td>
<td>16</td>
</tr>
<tr>
<td>Y206A/Y217A</td>
<td>2</td>
</tr>
<tr>
<td>RSV NC</td>
<td>0</td>
</tr>
</tbody>
</table>

a The binding affinities of single and double amino acid mutant proteins were examined by GEMSA with in vitro-transcribed full-length HTNV S-segment vRNA, quantitated by phosphorimaging, and compared to the wild-type HTNV N protein.
activity. Therefore, two double mutants were produced and examined by GEMSA (Fig. 5). One double mutant protein was made with the mutations Y178A/Y206A. The mutant protein Y178A maintained 100% signal intensity compared to the wt HTNV N protein. However, the double mutant showed only 5% of the RNA bound by the wt compared to 22% for the Y206A mutant alone (Fig. 5, compare lanes 4 and 5; Table 2). Strikingly, the Tyr-to-Ala changes at positions 206 and 217 showed almost complete inhibition, with only 2% of the signal present, as measured by phosphorimaging (Fig. 5, lane 7; Table 2). We have previously shown that the interaction between the HTNV N protein and the viral RNA is specific (32). As a control for specificity herein, the purified Rous sarcoma virus (RSV) nucleocapsid protein (NC) was included in the GEMSA to determine its ability to bind to HTNV vRNA. There was no gel shift with the RSV NC protein (Fig. 5, lane 9). In the same vein, the HTNV N protein did not bind to the packaging signal of the RSV RNA (data not shown). Finally, a GEMSA analysis of single Ala substitutions did not produce a band/gel shift when tested in the presence of a nonviral RNA substrate (data not shown).

**DISCUSSION**

The hantavirus N protein is instrumental for encapsidation of the viral RNAs for formation of the nucleocapsid and assembly of the virion. In addition, it has been suggested to play a role in the replication and transcription activities of the virus. Each of these activities requires an interaction with viral RNA. Previously, we mapped an RNA binding domain within the HTNV N protein to amino acids 175 to 217 (37). This region was mapped by its preferred interaction with full-length S-segment vRNA and oligoribonucleotides with the vRNA sequence. Herein, we refined the mapping of the RBD and identified the region spanning amino acids 194 to 204 as the minimal region for binding. GEMSA analysis showed that the chemical modification of Lys residues within the 175-429 region completely abolished the RNA binding activity (Fig. 3). In contrast, the chemical modification of Arg residues did not affect the RNA binding activity (Fig. 5). In support of these findings, two Ala substitutions, at Lys197 and Lys199 in the RBD, reduced RNA binding fivefold compared to that of the wt HTNV N protein (Table 3), while Arg substitutions reduced binding in peptides only twofold (Table 2). Three additional amino acids were mapped within the context of the full-length N protein, namely, Glu192, Tyr206, and Ser217 (Fig. 4). The double mutant protein, Y206A/S217A, showed almost complete inhibition of RNA binding (Fig. 5; Table 3). In conclusion, our work has demonstrated that specific amino acids located in the RBD (aa 175 to 217) contribute to RNA binding.

Studies with other RNA viruses have shown that there are specific RNA binding domains in their respective N proteins. In the mouse hepatitis virus nucleocapsid protein, a 55-amino-acid region from amino acids 177 to 231 bound vRNA with a dissociation constant of 32 nM (23). In Sindbis virus, a region in the nucleocapsid from residues 97 to 106 dictates specific encapsidation (25). In contrast, Elton et al. determined that multiple regions of the influenza virus nucleocapsid protein are essential for RNA binding (8). They showed by fluorescence spectroscopy that five tryptophan residues, one phenylalanine, and two arginines distributed throughout the protein are critical for high-affinity RNA binding (8). In comparison, our data implicate three residues, namely, a glutamic acid, a tyrosine, and a serine in the RBD, that may be essential for the affinity of the RNA-N protein interaction. Furthermore, the chemical modification of lysines, arginines, and cysteines showed that lysine residues scattered throughout the protein additionally contribute to the binding of the viral RNA, particularly in the region spanning aa 175 to 429. Also, lysine residues within the 1-70 region, predicted to be disordered and to have a propensity for induced folding, may also play a crucial role in RNA binding. Notably, Mir and Panganiban recently reported that the high-affinity binding of HTNV N to panhandle RNA relies upon a specific interaction between RNA and an RNA binding domain of trimeric N followed by a conformational change in the latter (21). The putative α-helical transition occurring within the 1-70 region of HTNV N after binding to RNA could be part of such a conformational change. Further studies are required to define the possible functional interactions of this region, which could be involved in subsequent assembly steps such as the multimerization of N and interactions of the viral genomic ends that can form panhandles.

In addition to the role the RBD plays in encapsidation of the RNA, studies aimed at determining if the RBD motif plays a role in N protein multimerization and subsequent assembly processes have been inconclusive. For example, in a study by Alfadhlí et al., the C-terminal regions of the Sin Nombre virus and Prospect Hill virus N proteins were required for N protein-N protein interactions in a yeast two-hybrid assay (2). Furthermore, in a recent study by Yoshimatsu et al., HTNV, Seoul virus, and Dobrava virus N protein multimerization was observed in a competitive enzyme-linked immunosorbent assay even when 49 amino acids were deleted from the N-terminal region (38). Additionally, using a yeast two-hybrid assay, they found that amino acids 100 to 125 and 404 to 429 were required for N protein multimerization. Collectively, these data indicate that at least two regions are required for multimerization of the hantavirus N protein. However, neither study speculated on the role of RNA in the assembly process. Overall, these data suggest that there may be an interdependence between nucleocapsid multimerization and N protein-RNA binding.

Our previous studies showed that the HTNV N protein-vRNA complexes are stable over a wide range of ionic strength conditions, suggesting that the N protein interaction with viral RNA relies on specific structural and/or sequence determinants in the 5′ end of the genome (32, 33). One hypothesis that Elton et al. proposed is that influenza virus RNA binding by the N protein is mediated by a combination of electrostatic and planar interactions through multiple regions of the protein (8). In contrast, we have recently found evidence from a thermodynamic analysis that the major contributions of the N protein-RNA interaction lay in van der Waals forces and hydrogen bonding (data not shown). It is intriguing that similar to the case for influenza virus (8), we found that positively charged amino acids, specifically lysine residues, distributed from amino acids 175 to 429 of the N protein were absolutely required for the N protein-RNA interaction, while the mutation of a negatively charged, an aromatic, and a hydroxyl-containing-side-chain amino acid in the RBD led to an 80% decrease
in RNA binding affinity. At a minimum, the evidence indicates that there are overlapping specificities between the lysine residues and the key amino acids in the RBD.

ACKNOWLEDGMENTS

This work was supported by a grant to C.J. from the Department of Defense (DAMD17-00-1-0513) and by the National Institutes of Health RISE grant GM61222-01.

We thank Salvador Cisneros, Samantha Claw, and Stephanie Chapman, interns in the NIH MBRS/RISE Program at New Mexico State University, for their technical assistance. We thank Maxine Linial at Fred Hutchinson Cancer Center for the Rous sarcoma virus nucleocapsid protein.

REFERENCES


