JOURNAL OF VIROLOGY, Oct. 2002, p. 10383–10392 Vol. 76, No. 20
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.20.10383–10392.2002


# A Structural Model of Pestivirus $E^{rns}$ Based on Disulfide Bond Connectivity and Homology Modeling Reveals an Extremely Rare Vicinal Disulfide

J. P. M. Langedijk,[1,2]* P. A. van Veelen,[3] W. M. M. Schaaper,[1] A. H. de Ru,[3] R. H. Meloen,[1] and M. M. Hulst[2]

*Pepscan Systems Inc., 8203 AB Lelystad,[1] Department of Mammalian Virology, Institute for Animal Science and Health (ID-Lelystad), 8200 AB Lelystad,[2] and Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, NL-2300 RC Leiden,[3] The Netherlands*

Received 30 April 2002/Accepted 2 July 2002

**$E^{rns}$ is a pestivirus envelope glycoprotein and is the only known viral surface protein with RNase activity. $E^{rns}$ is a disulfide-linked homodimer of 100 kDa; it is found on the surface of pestivirus-infected cells and is secreted into the medium. In this study, the disulfide arrangement of the nine cysteines present in the mature dimer was established by analysis of the proteolytically cleaved protein. Fragments were obtained after digestion with multiple proteolytic enzymes and subsequently analyzed by liquid chromatography-electrospray ionization mass spectrometry. The analysis demonstrates which cysteine is involved in dimerization and reveals an extremely rare vicinal disulfide bridge of unknown function. With the assistance of the disulfide arrangement, a three-dimensional model was built by homology modeling based on the alignment with members of the Rh/T2/S RNase family. Compared to these other RNase family members, $E^{rns}$ shows an N-terminal truncation, a large insertion of a cystine-rich region, and a C-terminal extension responsible for membrane translocation. The homology to mammalian RNase 6 supports a possible role of $E^{rns}$ in B-cell depletion.**

Classical swine fever virus (CSFV), bovine viral diarrhea virus (BVDV), and border disease virus belong to the genus *Pestivirus* of the *Flaviviridae* family (30), which infect pigs, cattle, and sheep. The disease is characterized by fever and hemorrhages and can run an acute or chronic course. Outbreaks of CSFV and BVDV infection can cause high economic losses. Like other members of the family, pestiviruses are plus-strand RNA viruses whose genome comprises one long open reading frame (5, 20, 30). Translation into a polyprotein is accompanied by processing into mature proteins. The structural proteins include a nucleocapsid protein, C, and three envelope glycoproteins, $E^{rns}$, E1, and E2 (26). The envelope glycoprotein $E^{rns}$ is a disulfide-linked homodimer of 100 kDa, and more than half of the molecular mass is contributed by carbohydrates (13, 22). It can be detected on the surface of pestivirus-infected cells and is secreted in the medium (22). $E^{rns}$ was able to bind many cell types (9), and binding of $E^{rns}$ is probably mediated by glycosaminoglycans (10, 11). Two sequence stretches of $E^{rns}$ show sequence homology to members of the $Rh/T_2/S$ RNase superfamily. A typical feature of this type of RNase is the relatively low base specificity and the high molecular weight (7). In line with this homology, $E^{rns}$ indeed contains RNase activity, which is unique for viral surface proteins (8, 23). Another unique feature of $E^{rns}$ is the dimeric state. Although multimerization is universal for viral surface proteins, for RNases it is described only for the cytotoxic bovine seminal RNase (BS-RNase), a member of the mammalian superfamily of "pancreatic-type" RNases (32). Like BS-RNase, $E^{rns}$ shows immunosuppressive activity since it induced apoptosis in concanavalin A-stimulated T cells of several species (3). In vivo studies in which the RNase activity of $E^{rns}$ in a recombinant virus was inactivated showed that the virus was attenuated and was not able to induce B-cell depletion. It was postulated that $E^{rns}$ might be responsible for the decrease in the B-cell count in infected pigs, most probably due to induction of apoptosis (18). However, the function of the RNase activity remains elusive. Because an extracellular protein with RNase activity most probably has an intracellular target, it was anticipated that $E^{rns}$ had some kind of way to enter the cell. Previously it was shown that $E^{rns}$ was indeed able to translocate into cells, and the translocation activity was mapped to the C-terminal domain (15).

Structural information would be very useful for a better understanding of this unique protein. Because crystallization of this heterogeneous glycoprotein was not successful (J. P. M. Langedijk, unpublished data), disulfide bonds were assigned and a structural model was built. The disulfide bonds are key structural elements that stabilize the functional native conformation and may also guide and test the hypothetical three-dimensional model. The redox state of the nine cysteine residues in the monomer is not known, but it is assumed that, as for other viral surface proteins, all cysteines are disulfide bonded.

* Corresponding author: Mailing address: Pepscan Systems Inc., Edelhertweg 15, P.O. Box 2098, 8203 AB Lelystad, The Netherlands. Phone: 31-320-237209. Fax: 31-320-238120. E-mail j.p.m.langedijk@id.wag-ur.nl.

## MATERIALS AND METHODS

$E^{rns}$ of CSFV strain C was expressed by a recombinant baculovirus in Sf21 cells as described previously (8, 10). In this study the $E^{rns}$ residues are numbered 1 to

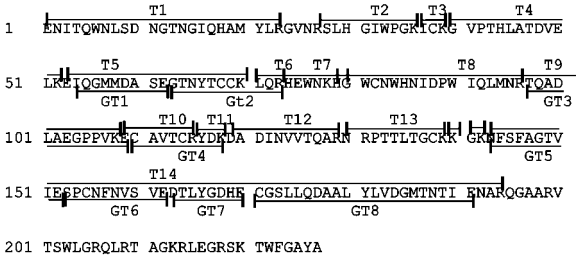


FIG. 1. Amino acid sequence of CSFV strain C $E^{rns}$. Residues 1 to 227 correspond to residues 268 to 494 of the CSFV polyprotein (20). Fragments of tryptic digests are coded T1 to T14; double digests are coded GT1 to GT8.

227, corresponding to amino acids 268 to 494 in the polyprotein of CSFV strain C (19).

**$E^{rns}$ production and purification.** Purified $E^{rns}$ was prepared as described previously (8). Briefly, Sf21 cells were infected with recombinant baculovirus expressing glycoprotein $E^{rns}$ of CSFV strain C. Infected cells were incubated for 96 h and then lysed in buffer containing 30 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, and 1% Nonidet P-40. The lysed cultures were centrifuged to remove cell debris, and the supernatant was stored at −20°C. $E^{rns}$ was purified by immunoaffinity chromatography with monoclonal antibody C5, directed against $E^{rns}$ of CSFV strain C. The concentrations of protein in the fractions containing RNase activity were determined by extrapolation of the absorption at 280 nm on a bovine serum albumin standard curve. RNase-specific activity was assayed as described previously (8) and expressed in units of absorption at 260 nm per minute per milligram. The recombinant protein reacted with MAb C5 and, like native $E^{rns}$, was shown to be a doublet of 42 to 46 kDa by immunoprecipitation. Its mobility was similar to that of $E^{rns}$ precipitated from SK6 cells infected with CSFV strain Brescia. The purity of the recombinant protein was more than 90% as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. For further purification, cation-exchange chromatography of $E^{rns}$ was performed on a Hitrap SP ion-exchange column (Pharmacia) and the $E^{rns}$ was eluted with NaCl. The specific RNase activity per milligram of the purified recombinant $E^{rns}$ was comparable to that of native $E^{rns}$ (23), indicating that the biological activity of $E^{rns}$ produced in insect cells is indistinguishable from that of native $E^{rns}$.

**Digestion.** $E^{rns}$ (200 μg in 0.1 M Tris-HCl [pH 7]–1.7 M urea–20 mM methylamine) was digested at 37°C with 4 μg of trypsin (modified, sequence grade [Boehringer Mannheim]). After 6 and 20 h, 1.75 μg of extra trypsin was added. After 24 h, 12 μg of trypsin inhibitor was added. The mixture was split into two aliquots. One aliquot was separated on a 4-kDa molecular filter (Nalge Nunc). The high-molecular-mass fraction was washed and transferred to 40 mM ammo-

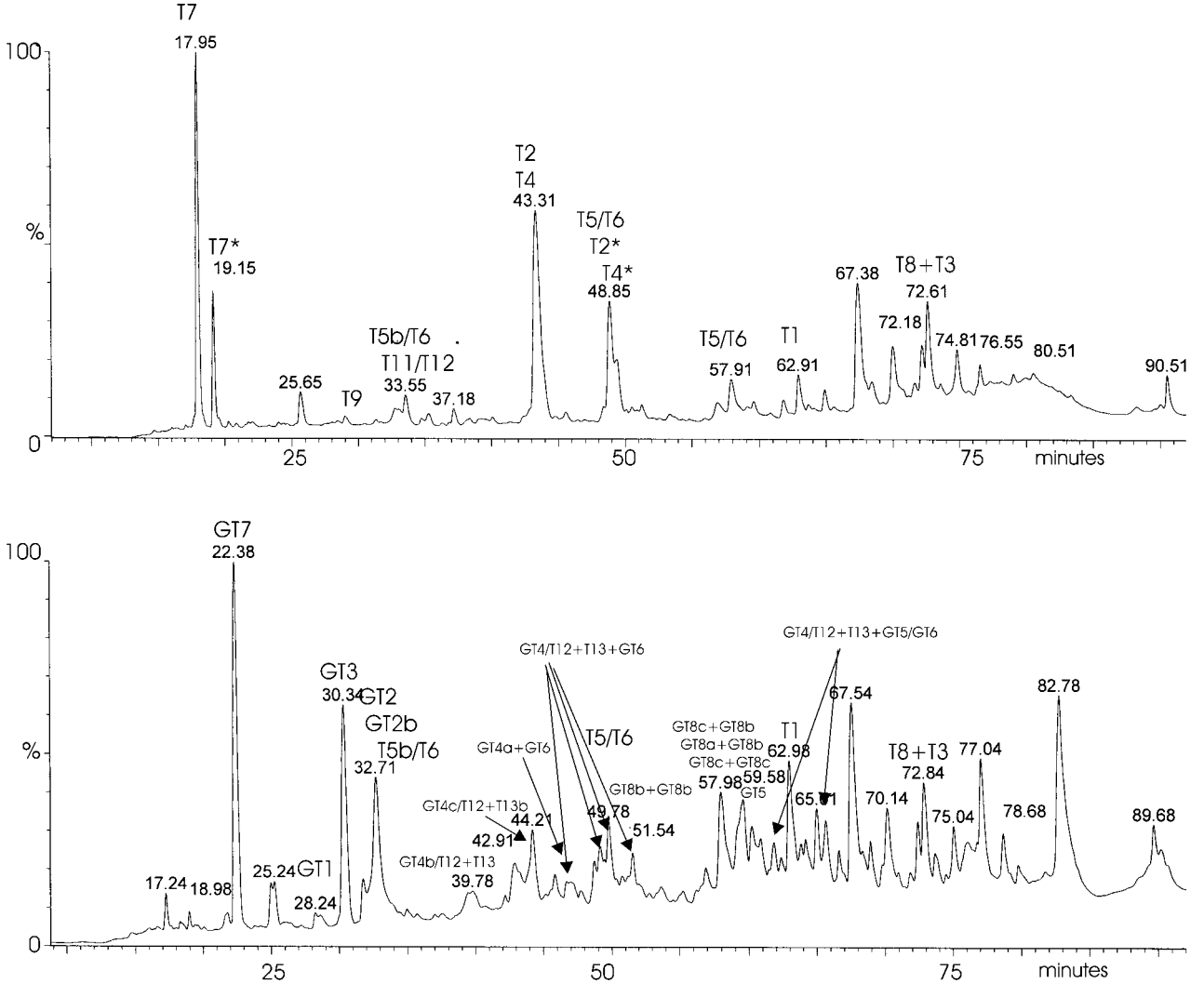


FIG. 2. HPLC chromatograms of trypsin-digested (top) and trypsin–Glu-C double-digested (bottom) $E^{rns}$, obtained by monitoring the UV absorbance at 215 nm. Fractions corresponding to the peaks were collected and analyzed. Detected fragments are indicated.

TABLE 1. Trypsin digestion

| Peptide | Retention time (min) | Sequence[a] | Measured mass (Da) | Theoretical mass (Da) | Measured reduced mass |
|---|---|---|---|---|---|
| T1 | 62.91 | **ENITQWNLSDNGTNGIQHAMYLR** | 2,820.3 | 2,674.3 | —[b] |
| T2 | 43.31 | SLBGIWPGK | 994.0 | 993.55 | — |
| T4 | 43.31 | GVPTHLWTDVELK | 1,378.5 | 1,378.75 | — |
| T5b/T6 | 32.80 | SEGTNYTCCKLQR | 1,500.5[f] | 1,499.67 | 1,502.5 |
| T5/T6 | 48.85 | EIQGMMDASEGTNYTCCKLQR | 2,375.8[f] | 2,375.02 | 2,377.8[d] |
| T7 | 17.95 | HEWNK | 712.5 | 712.34 | — |
| T8 + T3 | 72.61 | HGWCNWHNI**DPWIQLMNRTCK** ICK | 2,680.4[f] | 2,679.27 | 2,320.2[e] |
| T9 | 29 | TQADLAEGPPVK | 1,224.5 | 1,224.64 | NT[c] |
| T11/T12 | 33.55 | YDKDADINVVTQAR | 1,606.6 | 1,606.8 | NT |

[a] Several peptides were subjected to MS-MS. The elucidated sequence is underlined and bold.
[b] —, no difference between oxidized and reduced fragments.
[c] NT, not tested.
[d] Theoretical mass of reduced T5/T6, 2,377.0 Da.
[e] Theoretical mass of T8, 2,319.1 Da.
[f] Fragment is approximately 1 Da larger due to Asn to Asp modification during N deglycosylation.

nium acetate (pH 3.96). To this aliquot, urea (final concentration, 1 M), methylamine (final concentration, 20 mM) and 1 μg of endoproteinase Glu-C (sequencing grade; Boehringer Mannheim) were added. The mixture was incubated at 25°C for 20 h. Subsequently, $NH_4HCO_3$ was added to reach pH 7. To the other half of the trypsin digest, 15 μl of 300 mM Tris-HCl (pH 6.05) was added. Subsequently, 800 mU of *N*-glycosidase (E.C. 3.2.2.18; Boehringer Mannheim) was added to both aliquots for 2.5 h at 37°C to cleave N-linked glycosyl groups from the $E^{rns}$ protein fragments. Samples were stored at −20°C until analysis.

**HPLC purification.** The products were separated by reverse-phase high-performance liquid chromatography (HPLC) on an Alliance 2690 system (Waters, Milford, Mass.) equipped with a Delta-Pak $C_{18}$ column (300 Å, 5 μm, 2 by 150 mm [Waters]), using a gradient of water-acetonitrile containing 0.1% trifluoroacetic acid. The products were dried overnight using a speedvac. Samples were reduced by addition of 65 μl of 0.04 M $NH_4HCO_3$ and 0.6 μl of 1 M dithiothreitol (DTT). After incubation for 2 h at 20 °C, 4.3 μl of acetic acid (7.5%, vol/vol) was added and the samples were stored at −20°C until analysis.

**Mass spectrometry.** Electrospray ionization mass spectrometry (ESI-MS) was performed on a hybrid quadrupole time-of-flight mass spectrometer (Q-TOF; Micromass, Manchester, United Kingdom), equipped with an on-line nanoelectrospray interface (capillary tip, 20 μm [internal diameter] by 90 μm [outer diameter]) with an approximate flow rate of 250 nl/min. This flow was obtained by splitting of the 0.4-ml/min flow of a conventional high-pressure gradient system, using an Acurate flow splitter (LC Packings, Amsterdam, The Netherlands). Injections were done with a dedicated micro/nano-HPLC autosampler, the FAMOS (LC Packings, Amsterdam, The Netherlands), equipped with two extra valves for phase system-switching experiments. Digestion solutions were diluted in water-methanol-acetic acid (95:5:1, vol/vol/vol) and were trapped on the precolumn (MCA-300-05-C18; LC Packings) in water-methanol-acetic acid (95:5:1, vol/vol/vol). After washing of the precolumn to remove the buffers present in the digests, the trapped analytes were eluted with a step gradient from 70 to 90% water-methol-acetic acid (10:90:1, vol/vol/vol) in water-methanol-acetic acid (95:5:1, vol/vol/vol) in 10 min, with a flow of 250 nl/min. This low elution flow rate allows for a few additional MS-MS experiments if necessary during the same elution. Mass spectra were recorded from mass 50 to 2000 Da every second with a resolution of 5,000 full width half maximum. In MS-MS mode, ions were selected with a window of 2 Da with the first quadrupole and fragments were collected with high efficiency with the orthogonal time-of-flight mass spectrometer. The collision gas applied was argon (4 mPa) and the collision voltage was approximately 30 V. The MS and MS-MS spectra were interpreted with the help of the Biolynx software package supplied with the mass spectrometer.

## RESULTS

CSFV $E^{rns}$ (Fig. 1) expressed by insect cells was purified by immunoaffinity and ion-exchange chromatography, and 200 μg of purified $E^{rns}$ was digested with trypsin. Half of this digest was filtered, and subsequently, the larger fragments (>4 kDa) of this trypsin digest were double digested with endoproteinase Glu-C. The resulting fragments were N deglycosylated with PNGase F and separated by HPLC (Fig. 2), and fractions were subjected to ESI-MS analysis. The results of the analyses are summarized in Table 1 (tryptic digests) and Table 2 (double digests). Fragments were coded according to the nomenclature used in Fig. 1.

**Tryptic digests.** All measured tryptic digest fragments listed in Table 1 match the theoretical mass except for the T1 fragment (residues 1 to 23), which had an excess mass of 146 Da which could not be accounted for. MS-MS analysis of fragment T1 (*m*/*z* 941) identified the sequence $_2$DITQWNLSDDGTNG IQHAMYLR$_{23}$. This sequence implies that Asn2 and Asn11 may have been N glycosylated because N deglycosylation converts Asn residues into Asp residues. The N-terminal excess of 144 Da could be explained by two additional amino acids (Gly-Ser) that are introduced at the N terminus of $E^{rns}$ as a result of cloning of $E^{rns}$ into the *Bam*HI site behind the pseudorabies virus gX signal sequence in the baculovirus transfer vector (8). Not all theoretical trypsin cleavage sites were cleaved (e.g., Lys70 and Lys118), and due to the promiscuity of the enzyme, nontrypsin sites were also cleaved (Ala60). Fragments T2*, T4*, and T7* (Fig. 2, top) had masses of 1,037, 1,421.5 and 755.5 Da, respectively, which are 43 Da larger than the masses of T2, T4, and T7 (Table 1). This modification corresponds to an additional carbamyl group. Although methylamine was used as a scavenger during the trypsin digestion, the conditions (1.7 M urea [pH 7]) induced carbamylation of the N termini of several peptides, which caused different HPLC retention times. It is known that the cyanate ions present in the urea solution react with the α-amino group of peptides, forming carbamylated peptides (24). All fractions that contain the T5 fragment increased 2 Da in mass after reduction, which implies an adjacent disulfide bridge between cysteines 68 and 69 (Table 1). The T5 fragments may have been glycosylated at Asn65 because the measured mass of reduced T5 (2,377.8 Da) is 0.8 Da higher then the theoretical mass (2,377.0 Da). Because Asn65 is a potential glycosylation site, deglycosylation may account for the 1-Da difference because it converts the Asn into Asp. Reduction of the fraction at

TABLE 2. Trypsin–Glu-C double digestion

| Peptide | Retention time (min) | Sequence[a] | Measured mass (Da) | Theoretical mass (Da) | Measured reduced mass (Da) |
|---|---|---|---|---|---|
| T1 | 62.91 | ENITQWNLSDNGTNGIQHAMYLR | 2,820.0 | 2,674.3 | —[b] |
| GT1 | 28.24 | IQGMMDASE | 980.3 | 980.4 | — |
| GT2 | 32.71 | GTNYTCCKLQR | 1,284.5[e] | 1,283.6 | 1,286.5 |
| GT2b | 32.71 | **NYTCCKLQR** | 1,126.4[e] | 1,125.4 | 1,128.4 |
| T5b/T6 | 32.17 | SEGTNYTCCKLQR | 1,500.5[e] | 1,499.67 | 1,502.5 |
| T5/T6 | 49 | EIQGMMDASEGTNYTCCKLQR | 2,375.9[e] | 2,375.02 | 2,377.9 |
| GT3 | 30.3 | TQADLAEGPPVKE | 1,353.6 | 1,353.54 | — |
| GT5 | 59.58 | NFSFAGTVIE | 1,083.4 | 1,083.5 | — |
| GT7 | 22.38 | DTLYGDHE | 948.5 | 948.4 | — |
| GT8b + GT8b | 51.44 | CGSLLQCGSLLQ | 1,236.5 | 1,236.60 | 620 |
| GT8c + GT8b | 57 | C**GSLLQDA**<br>CGSLLQ | 1,422.6 | 1,422.66 | 619.3 + 805.4 |
| GT8a + GT8b | 57.98 | CGSLLQ<br>CGSLL | 1,108.5 | 1,108.54 | NT[c] |
| GT8c + GT8c | 58-61 | CGSLLQDA<br>CGSLLQDA | 1,608.6 | 1,608.72 | 805.3 |
| T8 + T3 | 72.61 | HGWCNWHNIDPWIQLMNR<br>ICK | 2,680.4[e] | 2,679.27 | NT |
| GT4/T12 + T13 + GT6 | 49.78 | CAVTCRYDKDADI**NVVTQAR**<br>NRPTTLTGCK<br>SPCNFN**VSVE** | 4,424.0[d,e] | 4,422.96[d] | 1,095.5 |
| GT4/T12 + T13 + GT6 | 47 | $+K_{(140)}$ | 4,549.2[d] | 4,549.96[d] | 1,095.56 |
| GT4/T12 + T13 + GT6 | 49 | $+K_{(140)}$ −TQAR | 4,092.6[d] | 4,094.67[d] | 1,095.36 |
| GT4/T12 + T13 + GT6 | 51.44 | −TQAR | 3,966.5[d] | 3,966.5[d] | 1,095.36 |
| GT4/T12 + T13 + GT5/GT6 | 66 | CAVTCRYDKDADI**NVVTQAR**<br>NRPTTLTGCK<br>NFS**FAGTVIES**PCNFNVSVE | 5,489.5[d,e] | 5,488.71[d] | NT |
| GT4/T12 + T13 + GT5/GT6 | 61.8 | $+_{(140)}KGK_{(142)}$ | 5,804[d,e] | 5,802.55[d] | NT |
| GT4 + T13 + GT5/GT6 | 66 | CAVTCR<br>NRPTTLTGCK<br>NFSFAGTVIESPCNFNVSVE | 3,898.8[d] | 3,899.45[d] | 651.15 |
| GT4a + GT6 | 47 | **CAV**<br>**SPCNFNVSVE** | 1,384.5[e] | 1,383.58 | 1,095.5 |
| GT4b/T12 + T13 | 39.78 | TCRYDKKDADINVV<br>NRPTTLTGCK | 2,599.2[d] | 2,599.95[d] | NT |

[a] Several peptides were subjected to MS-MS. The elucidated sequence is underlined and bold. − or + indicates identical sequence to previous without or with, respectively, the indicated residue(s).
[b] No difference between oxidized and reduced fragments.
[c] NT, not tested.
[d] Average mass.
[e] Fragment is approximately 1 Da larger due to Asn-to-Asp modification during N deglycosylation.

72.61 min resulted in the disappearance of the 2,680.4-Da fragment and the appearance of the reduced 2,320.2-Da fragment. This implies a disulfide bridge between cysteines 38 and 82 in fragments T3 (362.2 Da) and T8 (2,320.2 Da). The measured mass of T8 is 1.1 Da higher than the theoretical mass. Because Asn95 is a potential glycosylation site, deglycosylation may account for the 1-Da difference.

Only N-terminal fragments were identified after trypsin digestion. Due to the lack of trypsin cleavage sites and the possible cystine connectivity, no C-terminal fragments could be detected. Therefore, the trypsin digest was digested with endoproteinase Glu-C, as described in Materials and Methods.

**Trypsin–Glu-C double digests.** All trypsin–Glu-C double-digest fragments listed in Table 2 match the theoretical mass except for the T1 fragment (see above). As shown for the trypsin digests, the fragments that contained cysteines 68 and 69 (fragments GT2 and T5) showed a mass increase of 2 Da after reduction, implying the presence of an adjacent disulfide bridge between cysteines 68 and 69 (Fig. 3 and 4). MS-MS analysis of fragment GT2b (*m*/*z* 565) confirmed the sequence DYTCCQKLQR and also confirmed the Asn65-to-Asp65 conversion (Fig. 3c). In the fractions that eluted between 51 and 61 min, subfragments of GT8 were identified, which appeared to be disulfide linked to other GT8 subfragments (Fig. 5; Table 2). In most cases the reduced fragments could be detected. For the fraction at 57 min, the peptide sequence was fully confirmed by MS/MS analysis of fragment GT8c (*m*/*z* 805) (Table 2; Fig. 5c). Therefore, cysteines at position 171 form the intermolecular disulfide bridge, which connects two $E^{rns}$ monomers.

The remaining four cysteines which would probably form the remaining two disulfide bridges were difficult to determine. Detailed analysis of all the mass data and MS/MS data revealed that the two remaining disulfide bridges were contained in high-molecular-mass fragments (3,899 to 5,804 Da) in fractions that eluted between 47 and 66 min. Using the combined masses of the remaining fragments and the MS/MS sequence data, it could be deduced that the high-molecular-mass fragments contained smaller fragments corresponding to a combination of GT4/T12, T13, GT5, and GT6 (Fig. 4; Table 2). After reduction, only the GT6 fragment could be detected (Fig. 4b,

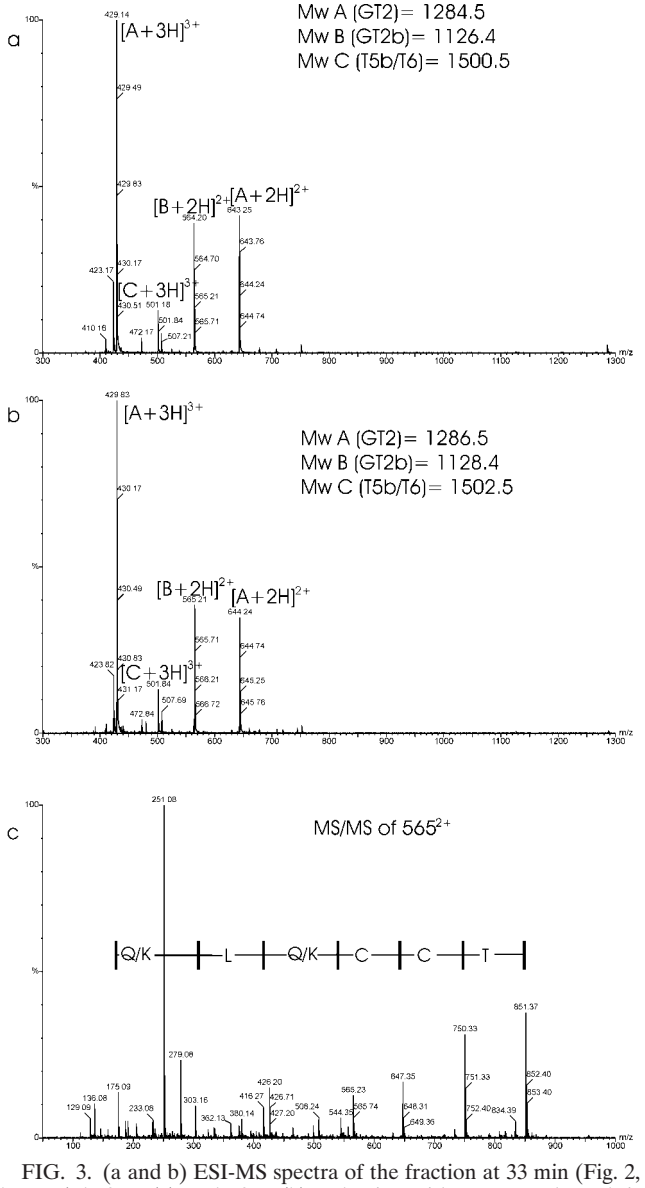


FIG. 3. (a and b) ESI-MS spectra of the fraction at 33 min (Fig. 2, bottom) before (a) and after (b) reduction with DTT. Each peak is labeled with an *m/z* ratio and a charge state designation. (c) Product ion mass spectrum of the GT2b peptide at *m/z* 565 (residues DYTC-CKLQR). The spectrum displays two series of peaks, the so-called b ions and y ions, that allow the sequence to be elucidated. A part of the y-ion series consists of the masses 851.4, 750.3, 647.4, 544.4, 416.3, 303.2, and 175.9, which correspond to the partial sequence T-C-C-(Q/K)-L-(Q/K)-R. A part of the b-ion series consists of the masses 279.08 and 380.14, which correspond to the partial sequence DY. Mw, molecular mass.

1,095 Da). The GT6 fragment may have been glycosylated at Asn158 because the measured mass of reduced GT6 (1,095.5 Da) is 1 Da higher than the theoretical mass (1,094.5 Da). MS-MS analysis confirmed the conversion of Asn to Asp (Table 2).

Using MS-MS analysis, we found more proof for the identity of the smaller fragments contained in the high-molecular-mass fragment (Table 2). MS-MS analysis revealed the sequences $_{123}$NVVTQAR$_{129}$ and $_{159}$VSVE$_{162}$ in a 4,549-Da fragment and the sequences $_{123}$NVVTQAR$_{129}$ and $_{146}$FAGTVIES$_{153}$ in a 5,489-Da fragment. The identity of the high-molecular-mass fragment indicated that cysteines 110 and 114 are not connected by a disulfide bridge but that these residues must be connected by a disulfide bridge to cysteines 138 and 155. This leaves only two possibilities for connectivity. Unfortunately, there is no cleavage site for trypsin or Glu-C between cysteines 110 and 114 (residues CAVTC). However, trypsin in particular can be promiscuous under the described conditions. Tables 1 and 2 show several examples of missed cleavages and also of fragments which have been cleaved after residue Ala, Leu, or Thr (fragments T5b, GT2b, and GT8a,c). Indeed, promiscuous cleavage between cysteines 110 and 114 could also be found in the fragments GT4a + GT6 and GT4b/T12 + T13 (Table 2). At 47 min in the chromatogram, a fragment of 1,384.5 Da was detected, and after reduction the GT6 fragment was detected. The peptide sequence was also fully confirmed by MS/MS (Table 2). Therefore, cysteines 110 and 155 must be connected by a disulfide bridge. This leaves only cysteines 114 and 138 that must form the final disulfide bridge. Proof of this connectivity was found in the sample at retention time 39.78 min, which has a mass that corresponds to GT4b/T12 + T13 (Table 2).

**Modeling.** Two stretches of *Pestivirus* E$^{rns}$ show sequence homology to members of the Rh/T$_2$/S RNase superfamily (7). The crystal structure of the family members RNase Rh, RNase MC, RNase S3, and RNase LE have been determined (14, 17, 21, 25), and the three-dimensional (3D) structures confirmed that both stretches with sequence homology to E$^{rns}$ constitute the active site of the RNase. Apart from the two stretches of sequence homology, further homology in the rest of the protein was not apparent. Despite the low sequence homology, a complete alignment was constructed using different types of scoring matrices and multiple sequence alignment of a large set of RNase sequences. A satisfactory alignment was not possible using alignment software with any parameter setting. Therefore, part of the alignment was edited manually. For parts with low sequence homology, the alignment was guided by the secondary-structure prediction of the PHD software (B. Rost and C. Sander, Nature **360:**540, 1992).

The multiple-sequence alignment is shown in Fig. 6. The parts of the E$^{rns}$ sequences with low homology or conflicting secondary structure predictions are shown in bold type.

A 3D structure of CSFV E$^{rns}$, strain Brescia (Fig. 7), was built by homology modeling based on the alignment of Fig. 6, using the RNase Rh crystal structure as a template. The model was built using the biopolymer package of Sybyl. Loop searches were performed for residues 37 to 38 (loop 2), 47 to 49 (loop 3), 63 to 66 (loop 4), 83 to 85 (loop 5), 93 to 95 (loop 6), 108 to 140 (loop 7), 153 to 163 (loop 8), and 171 to 174 (loop 9). The C-terminal 37 residues were modeled on the backbone of the L3 loop of restrictocin (31), which shows some homology (15). For all loops except loops 1 and 7, database structures were chosen with the highest score according to the Loop_search program. For loop 1, the homologous loop of the crystal structure of RNase MC was used (21). For loop 7, a loop was chosen that did not clash with the rest of the structure and did not extend too far from the protein surface and could accommodate the disulfide connectivity. The amino acid identity to the fungal RNase Rh is 15%. When the two stretches of

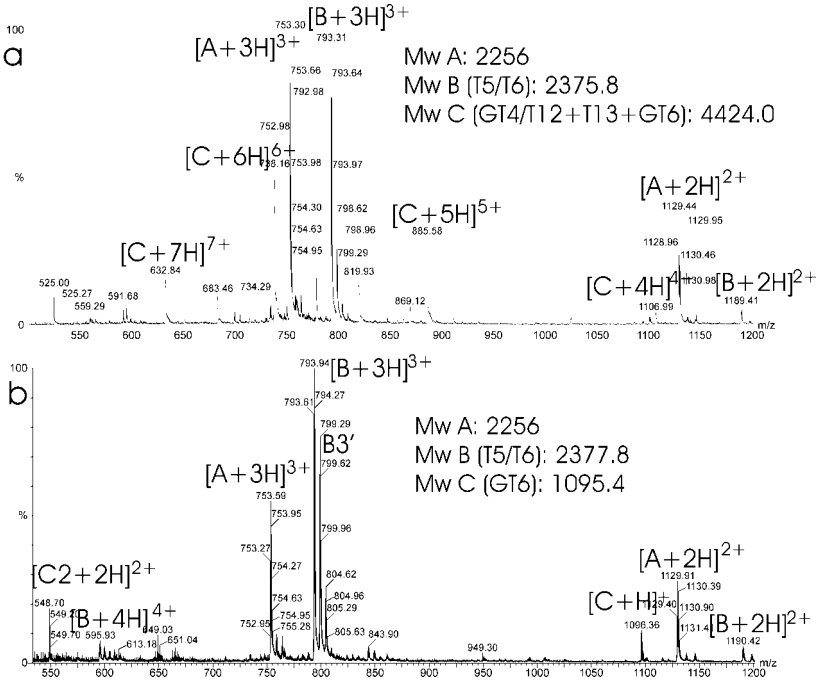


FIG. 4. ESI-MS spectra of the fraction at 49.5 min (Fig. 2, bottom) before (a) and after (b) reduction with DTT. Each peak is labeled with an *m/z* ratio and a charge state designation. Calculated masses (A to D) are indicated in the figure. B corresponds to fragment T5T6. C corresponds to a disulfide-containing fragment that contains fragment GT6. Mw, molecular mass.

homologous active-site regions are excluded, the identity is only 11%. The 3D structures, however, may be similar according to the model, specifically, the N-terminal part with the active site (blue in Fig. 7) and strand 4 (yellow in Fig. 7). The N-terminal part (from S1 to the loop between H4 and H5) is most highly conserved between RNase Rh/T2/S family members, between Rh/T2/S RNase family members and pestivirus $E^{rns}$, and between pestivirus $E^{rns}$ sequences. Compared with the 3D structure of RNase Rh, which was used as a template for modeling, $E^{rns}$ shows a considerable N-terminal truncation of 27 residues. Other deletions are found between S2 and H2 (9 residues) and between H4 and H5 (13 residues). An insertion is located between S1 and H1 (10 residues) a large extra cysteine-rich domain is found between H6 and H7 (21 residues), and an insertion of 7 residues is found between H7 and S3. $E^{rns}$ is able to bind glycosaminoglycans such as heparan sulfate (10, 11). When the electrostatic potential of the $E^{rns}$ model is calculated, several positively charged regions occur on the protein surface that may be involved in heparan sulfate interaction. The C-terminal region is especially highly charged, and peptides corresponding to this region interact with dextran sulfate columns (Langedijk, unpublished); Arg209 in this region was shown to be involved in heparan sulfate binding (10). Depending on the exact location of the C-terminal domain relative to the rest of the protein, it may form a continuous positively charged region with the basic helix H4. Another positively charged region is located between Cys114 and helix 7.

## DISCUSSION

The disulfide bonds of the CSFV $E^{rns}$ glycoprotein have been determined by ESI-MS and MS-MS after trypsin and Glu-C digestion at pH 7 and pH 4, respectively. Neutral pH for trypsin digestion was used to avoid disulfide interchange. In this study, each of the nine cysteines was found in a unique disulfide bond, which indicates that no disulfide rearrangement had occurred. The $E^{rns}$ dimer contains four intramolecular disulfides and one intermolecular disulfide. Despite the carbamylation, the partial digestion, the promiscuous digestion, and the lack of a cleavage site between cysteines 110 and 114, the study has established the complete location of the disulfide bridges in the dimer of CSFV $E^{rns}$. Cys171 is involved in $E^{rns}$ dimerization since it forms a disulfide bridge with Cys171′ of the other monomer in the $E^{rns}$ dimer. Interestingly, Cys171 is the only cysteine that is not completely conserved in $E^{rns}$. Several strains of BVDV type 1 and CSFV have no cysteine at position 171 and contain only eight cysteines (accession no. O11993, O11994, Q91WA6, P19711, and Q98Y26). Therefore, $E^{rns}$ monomers of these wild-type strains cannot dimerize just by forming a disulfide bridge between Cys171 residues. Perhaps other contacts or a domain swap are involved in dimerization. It would be interesting to study the occurrence and stability of $E^{rns}$ dimers in these strains. The other disulfide bridges are located between Cys38 and Cys82, which is conserved between all RNases of the Rh/T2/S family, Cys110 and Cys155, and Cys114 and Cys138, and a very rare vicinal disulfide is located between the adjacent cysteines 68 and 69 (Fig. 7 and 8). Because a disulfide between two adjacent cysteines cannot play a long-range architectural role, it is likely that the disulfide serves another function. This extremely rare vicinal disulfide is observed in only three other proteins: the α subunit of acetylcholine receptor (αAChR) (12), methanol dehydrogenase (MDH) (2) and the Janus-faced atracotoxins (J-ACTXs)

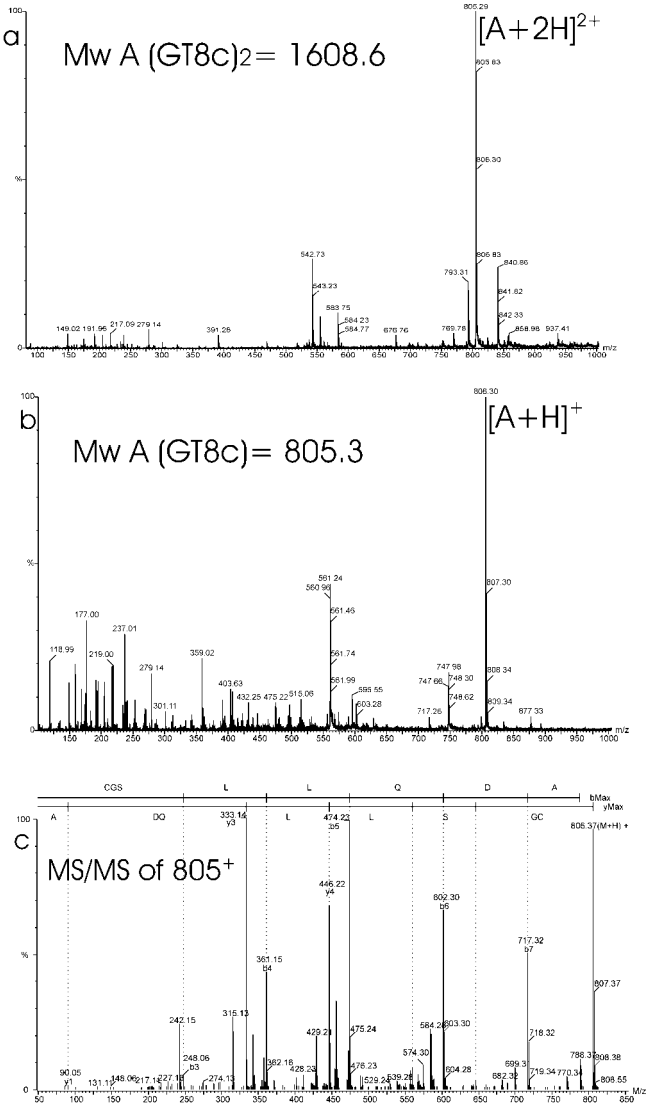


FIG. 5. (a and b) ESI-MS spectra of the fraction at 58.6 min (Fig. 2, bottom) before (a) and after (b) reduction with DTT. Each peak is labeled with an *m/z* ratio and a charge state designation. A corresponds to fragment GT8c+GT8c. (c) Product ion mass spectrum of the peptide GT8c at *m/z* 805 (residues CGSLLQDA). The spectrum displays two series of peaks, the b ions and y ions, that allow the sequence to be elucidated. A part of the y-ion series consists of the masses 646.3, 559.3, 446.2, 333.1, 205.1, and 90.1, which correspond to the partial sequence S-L-L-(Q/K)-D-A. A part of the b-ion series consists of the masses 248.1, 361.2, 474.3, 602.3, and 717.3, which correspond to the partial sequence G-S-L-L-(Q/K)-D. Mw, molecular mass.

(29). In αAchR, the disulfide is involved in acetylcholine binding; in MDH, the enzymatic activity is abolished when the disulfide at the enzyme active site is reduced; and in J-ACTX, the disulfide is critical for insecticidal activity and may be directly involved in interactions with the target molecule. Vicinal disulfides are readily reduced in αAchR and MDH; because of possible strained cyclic geometry, the disulfide may be highly reactive.

The function of E^rns and the importance of its dimeric nature are not understood. The only other known dimeric RNase is BS-RNase, which belongs to the monomeric "pancreatic-type" family. The dimeric nature is crucial for the cytotoxic and immunosuppressive activity of BS-RNase. BS-RNase has evolved from pancreatic RNase A, and the evolution of the dimeric state was accompanied by the acquisition of two adjacent cysteine residues at positions 31 and 32 at the end of a helix. These cysteines are involved in dimer formation using either the predominant "antiparallel" 31–32′/32–31′ connectivity with an identical N-terminal domain swap between monomers or "parallel" 31–31′/32–32′ connectivity without an N-terminal swap (6). However, under physiological conditions, 50% of the BS-RNase appears to be conjugated to other proteins, presumably through intermolecular disulfide bridges (4). Therefore, perhaps BS-RNase also contains an adjacent cystine bridge between Cys31 and Cys32, which could rapidly form mixed disulfides with another protein or a dimer when it encounters another monomer (4). Thus, although E^rns and BS-RNase are from different RNase families, they share two unique features: the dimeric nature and probably the vicinal disulfide. For E^rns, it may also imply that the connectivity of the adjacent cysteines 68 and 69 may be dependent on the environment of the protein and may dimerize through the cysteines or couple to other unknown proteins under physiological conditions. If dimerization via cysteines 68 and 69 can occur, it is questionable if this could occur together with dimerization via cysteine 171. According to the model (Fig. 7), cysteines 68, 69, and 171 are on the same side of the protein but the helix and loop between S1 and S2 obstruct a connection of all three cysteines. As mentioned above, a disulfide bridge between Cys171 and Cys171' is not essential because variant strains are described without a cysteine at position 171. Recently, Cys171 of recombinant CSFV was mutated, resulting in a viable virus with an E^rns mutant that was indeed unable to form an intermonomer disulfide bridge. Therefore, there is no proof that cysteines 68 and 69 compensate for the loss of the intermonomer disulfide in this mutant (H. G. P. van Gennip, unpublished data), and in contrast to cysteines 31 and 32 in BS-RNase, there is no proof for Erns dimerization through cysteines 68 and 69.

After inspection of the multiple sequence alignment, some major dissimilarities between the sequences of pestivirus E^rns and the other RNases can be observed (Fig. 6). The sequence of pestivirus E^rns is by far the most dissimilar to the known plant, fungal, bacterial, and mammalian Rh/T2/S RNases. If pestiviruses acquired the mammalian cellular RNase gene by RNA recombination, it has diversified considerably outside the active-site regions, because the sequence is very dissimilar from that of human RNase 6 (Fig. 6) (27) and the homologous mouse, pig, bovine, or chicken RNase (data not shown). The function of the mammalian RNase 6 may be helpful for the elucidation of the function of E^rns and vice versa. Interestingly, RNase 6 may be involved in the suppression of B-cell non-Hodgkin's lymphomas and acute lymphoblastic leukemias (1), while E^rns may be involved in the depletion of B lymphocytes (18), which indicates that the functions of the two proteins may be more similar than the low sequence homology would suggest.

The most important differences between E^rns and the other members of the Rh/T2/S RNase family is the truncated N terminus, the large insertion of a cysteine-rich domain between

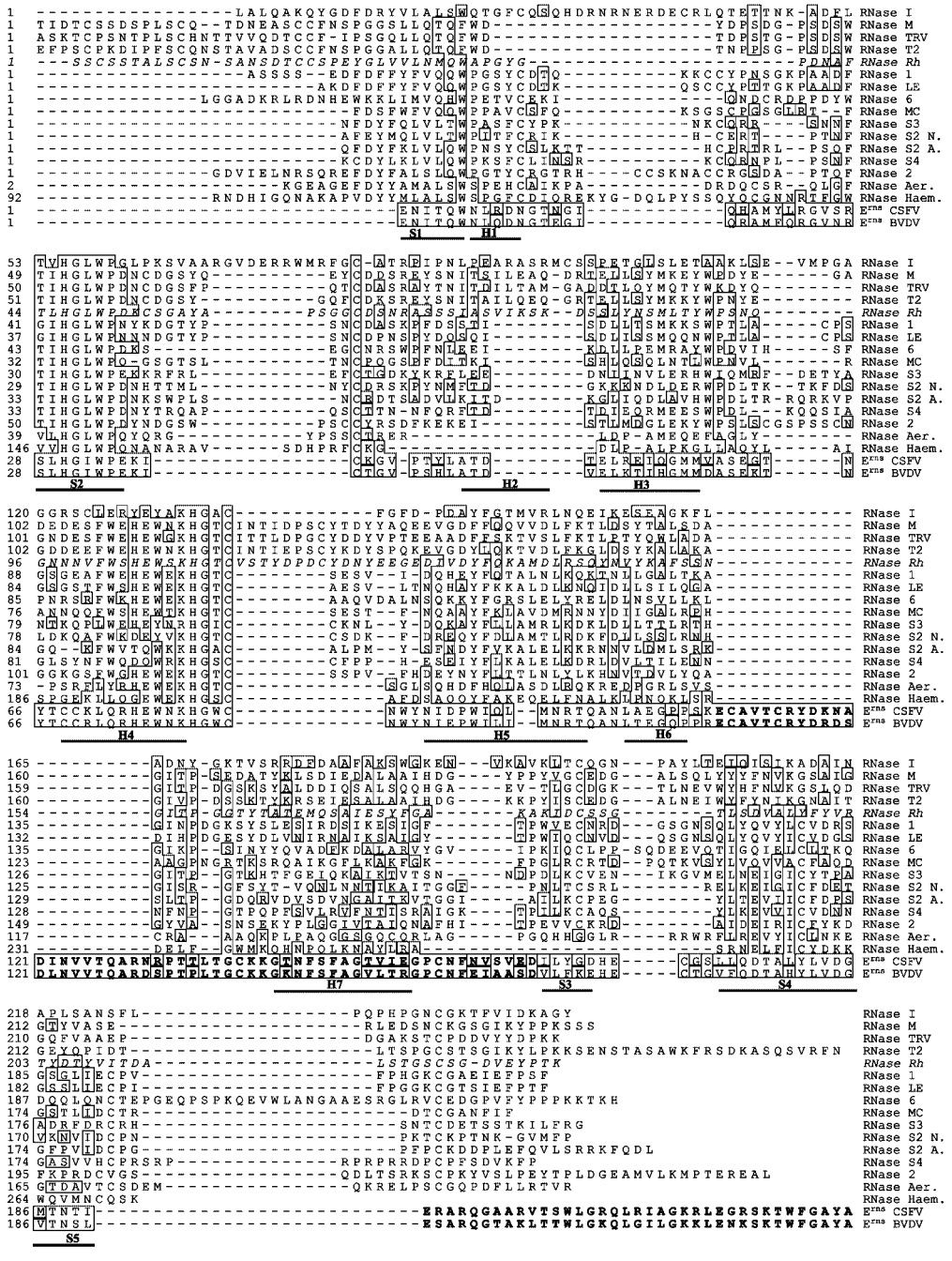

FIG. 6. Alignment of several RNases belonging to the RNase Rh/T2/S family. RNase I, *Escherichia coli* (accession no. P21338); RNase M, *Aspergillus saitoi* (P19791); RNase TRV, *Trichoderma viride* (P24657); RNase T2, *Aspergillus oryzae* (P10281); RNase Rh, *Rhizopus niveus* (P08056); RNase 1, *Arabidopsis thaliana* (P42813); RNase LE, *Lycopersicon esculentum* (P80022); RNase 6, *Homo sapiens* (NP003721); RNase MC, *Momordica charantia* (P23540); RNase S3, *Petunia* (Q40875); RNase S2, N., *Nicotiana alata* (P04007); RNase S2 A., *Antirrhinum hispanicum* (Q38716); RNase S4, *Antirrhinum hispanicum* (Q38717); RNase 2, *Arabidopsis thaliana* (P42814); RNase Aer., *Aeromonas hydrophila* (Q07465); Probable RNase Haem., *Haemophilus influenzae* (P44012); $E^{rns}$ of CSFV, strain Brescia (P21530); $E^{rns}$ of BVDV, strain SD-1 (Q01499). The crystal structure of RNase Rh (italic sequence) was used as a template for the 3D model of CSFV $E^{rns}$. Secondary-structure elements are indicated by bars and are marked S1 to S5 (strands) and H1 to H7 (helices). Secondary structures are based on crystal structure of RNase Rh except for helix H1, found in RNase MC, RNase S3, and RNase LE. Residues within 1 distance unit from CSFV are boxed. Gaps are indicated by dashes. Sequences of pestivirus $E^{rns}$ that could not be aligned are shown in bold type.

helix 6 and helix 7, and an elongated and very dissimilar C-terminal region (shown in red in Fig. 7). The structural differences of $E^{rns}$ from the RNase Rh template structure are the loss of several clustered loops or domains on one side of the protein (compared with the RNase Rh template protein) and the addition of large clustered loops or domains on the opposite side of the protein. The exact structure of the large insertions cannot be predicted. Specifically, the cysteine-rich do-

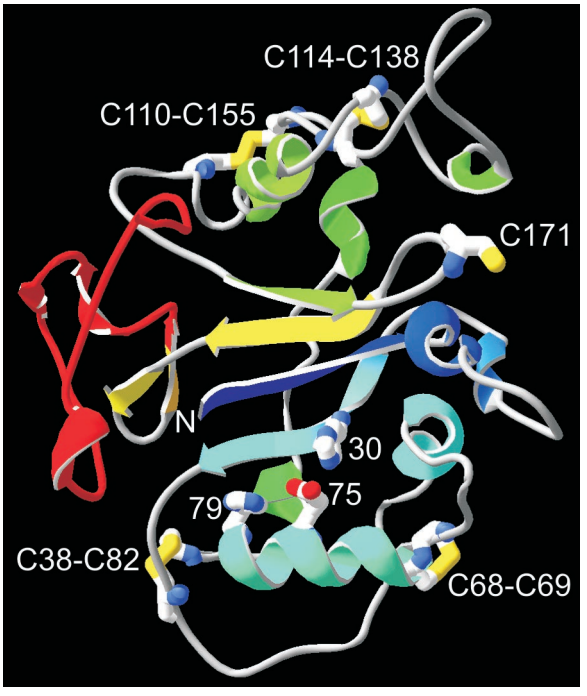


FIG. 7. 3D model of CSFV $E^{rns}$. Disulfides, unpaired Cys171, and active-site residues His30, Glu75, and His79 are shown as stick models. Secondary-structure elements are colored blue to red from the N to the C terminus. The C-terminal domain responsible for translocation is colored red.

main (residues 107 to 141) and the C-terminal 37 residues should be considered a rough approximation. The cysteine-rich domain is constrained by its N and C termini and the two disulfides, but the C-terminal region can be folded to different parts of the protein. The C-terminal region contains a large number of positive charges, has a high score for amphipathic helicity, and is an important antigenic site for all pestivirus $E^{rns}$ proteins (16). According to the alignment, the 37 C-terminal residues form a separate domain that shows more homology to the L3 loop of ribotoxins type II and to membrane-active peptides such as magainins (15). Previously it was shown that the C-terminal domain of $E^{rns}$ and the homologous L3 loop of ribotoxins were responsible for protein translocation across the plasma membrane (15). Therefore, the structure of the C-terminal region may be metastable and may change its conformation when it contacts heparan sulfate or the plasma membrane.

The biochemical analysis did not address the position and nature of the carbohydrates, but the lower mass of four fragments and MS-MS analysis indicate that Asn2, Asn11, Asn65, Asn95, and Asn158 may have been N glycosylated. Deglycosylation converts an Asn residue into an Asp residue, resulting in the loss of 1 Da. Asn2 is close to the active site, but the side chain (and the possible carbohydrate) is directed away from the active site. It is not clear from the model if a bulky carbohydrate at Asn2 can be accommodated in the 3D model. All other positions at which Asn-to-Asp conversion occurred and may have been N glycosylated are in good agreement with the model. Asn11 is located on the exposed loop between helix 1 and strand 2. Asn65 is located on the exposed loop between helix 3 and helix 4; Asn95 is located on "disrupted" helix 5, which runs under the β-sheet; and Asn158 is located on the loop between helix 7 and strand 3 (not present in BVDV). In the recently published crystal structure of RNase S3, N-linked glycans were observed at the same location as $E^{rns}$ Asn11 and Asn95 (17). The exposed C-terminal region of $E^{rns}$ is completely devoid of potential glycosylation sites. An interesting sequence motif ($_{81}$WXXW$_{84}$) is located between helix 4 and helix 5. Recently, this motif has been recognized to be a potential recognition site for C mannosylation, a novel type of glycosylation that involves the attachment of an α-mannosyl residue to the C-2 atom of the first tryptophan. Little is known about the role of C mannosylation. Structural studies of RNase 2 show that the main structural role of the mannose residue is stabilization of a loop and the ability to keep tryptophan in a specific orientation (28). It is not known whether Sf21 cells can carry out C mannosylation.

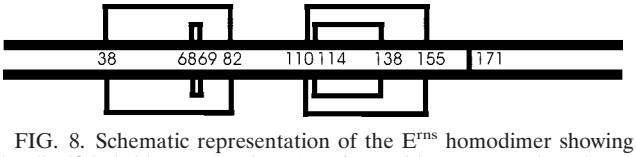


FIG. 8. Schematic representation of the $E^{rns}$ homodimer showing the disulfide bridge connection. Cysteine residues are numbered.

The combination of MS of peptide fragments and homology modeling has allowed us to assign disulfide bonds and propose a 3D model for $E^{rns}$. $E^{rns}$ is quite atypical compared with the other Rh/T2/S RNases. The $E^{rns}$ sequence is by far the most dissimilar; the protein is homodimeric, contains a translocation domain, and contains a motif for potential C mannosylation; the structure contains two disulfide bridges in a unique cysteine-rich region; and the protein contains a very rare vicinal disulfide bridge. This list of exciting features will help future functional studies on this secreted RNase.

## REFERENCES

1. **Acquati, F., C. Morelli, R. Cinquetti, M. G. Bianchi, D. Porrini, L. Varesco, V. Gismondi, R. Rocchetti, S. Talevi, L. Possati, C. Magnanini, M. G. Tibiletti, B. Bernasconi, M. G. Daidone, V. Shridhar, D. I. Smith, M. Negrini, G. Barbanti-Brodano, and R. Taramelli.** 2001. Cloning and characterization of a senescence inducing and class II tumor suppressor gene in ovarian carcinoma at chromosome region 6q27. Oncogene **20:**980–988.
2. **Blake, C. C., M. Ghosh, K. Harlos, A. Avezoux, and C. Anthony.** 1994. The active site of methanol dehydrogenase contains a disulphide bridge between adjacent cysteine residues. Nat. Struct. Biol. **1:**102–105.
3. **Bruschke, C. J., R. J. Moormann, J. T. van Oirschot, and P. A. van Rijn.** 1997. A subunit vaccine based on glycoprotein E2 of bovine virus diarrhea virus induces fetal protection in sheep against homologous challenge. Vaccine **15:**1940–1945.
4. **Ciglic, M. I., P. J. Jackson, S. A. Raillard, M. Haugg, T. M. Jermann, J. G. Opitz, N. Trabesinger-Ruf, and S. A. Benner.** 1998. Origin of dimeric structure in the ribonuclease superfamily. Biochemistry **37:**4008–4022.
5. **Colett, M. S., R. Larson, C. Gold, D. Strick, D. K. Anderson, and A. F. Purchio.** 1988. Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhea virus. Virology **165:**191–199.
6. **D'Alessio, G.** 1999. Evolution of oligomeric proteins. The unusual case of a dimeric ribonuclease. Eur. J. Biochem. **266:**699–708.
7. **Horiuchi, H., K. Yanai, M. Takagi, K. Yano, E. Wakabayashi, A. Sanda, S. Mine, K. Ohgi, and M. Irie.** 1988. Primary structure of a base non-specific ribonuclease from *Rhizopus niveus.* J. Biochem. (Tokyo) **103:**408–418.
8. **Hulst, M. M., G. Himes, E. Newbigin, and R. J. Moormann.** 1994. Glycoprotein E2 of classical swine fever virus: expression in insect cells and identification as a ribonuclease. Virology **200:**558–565.
9. **Hulst, M. M., and R. J. Moormann.** 1997. Inhibition of pestivirus infection in cell culture by envelope proteins E(rns) and E2 of classical swine fever virus: E(rns) and E2 interact with different receptors. J. Gen. Virol. **78:**2779–2787.
10. **Hulst, M. M., H. G. P. van Gennip, and R. J. M. Moormann.** 2000. Passage

of classical swine fever virus in cultured swine kidney cells selects virus variants that bind to heparan sulfate due to a single amino acid change in envelope protein $E^{rns}$. J. Virol. **74:**9553–9561.
11. **Iqbal, M., H. Flick-Smith, and J. W. McCauley.** 2000. Interactions of bovine viral diarrhoea virus glycoprotein E(rns) with cell surface glycosaminoglycans. J. Gen. Virol. **81:**451–459.
12. **Kao, P. N., and A. Karlin.** 1986. Acetylcholine receptor binding site contains a disulfide cross-link between adjacent half-cystinyl residues. J. Biol. Chem. **261:**8085–8088.
13. **Konig, M., T. Lengsfeld, T. Pauly, R. Stark, and H. J. Thiel.** 1995. Classical swine fever virus: independent induction of protective immunity by two structural glycoproteins. J. Virol. **69:**6479–6486.
14. **Kurihara, H., T. Nonaka, Y. Mitsui, K. Ohgi, M. Irie, and K. T. Nakamura.** 1996. The crystal structure of ribonuclease Rh from Rhizopus niveus at 2.0 A resolution. J. Mol. Biol. **255:**310–320.
15. **Langedijk, J. P.** 2002. Translocation activity of C-terminal domain of pestivirus Erns and ribotoxin L3 loop. J. Biol. Chem. **277:**5308–5314.
16. **Langedijk, J. P. M., W. G. J. Middel, R. H. Meloen, J. A. Kramps, and J. A. de Smit.** 2001. Enzyme-linked immunosorbent assay using a virus type-specific peptide based on a subdomain of envelope protein $E^{rns}$ for serologic diagnosis of pestivirus infections in swine. J. Clin. Microbiol. **39:**906–912.
17. **Matsuura, T., H. Sakai, M. Unno, K. Ida, M. Sato, F. Sakiyama, and S. Norioka.** 2001. Crystal structure at 1.5-A resolution of *Pyrus pyrifolia* pistil ribonuclease responsible for gametophytic self-incompatibility. J. Biol. Chem. **276:**45261–45269.
18. **Meyers, G., A. Saalmüller, and M. Büttner.** 1999. Mutations abrogating the RNase activity in glycoprotein $E^{rns}$ of the pestivirus classical swine fever virus lead to virus attenuation. J. Virol. **73:**10224–10235.
19. **Moormann, R. J., H. G. van Gennip, G. K. Miedema, M. M. Hulst, and P. A. van Rijn.** 1996. Infectious RNA transcribed from an engineered full-length cDNA template of the genome of a pestivirus. J. Virol. **70:**763–770.
20. **Moormann, R. J. M., A. Bouma, J. H. Kramps, C. Terpstra, and A. J. De Smit.** 2000. Development of a classical swine fever subunit marker vaccine and companion diagnostic test. Vet. Microbiol. **73:**209–219.
21. **Nakagawa, A., I. Tanaka, R. Sakai, T. Nakashima, G. Funatsu, and M. Kimura.** 1999. Crystal structure of a ribonuclease from the seeds of bitter gourd (*Momordica charantia*) at 1.75 A resolution. Biochim. Biophys. Acta **1433:**253–260.
22. **Rumenapf, T., G. Unger, J. H. Strauss, and H. J. Thiel.** 1993. Processing of the envelope glycoproteins of pestiviruses. J. Virol. **67:**3288–3294.
23. **Schneider, R., G. Unger, R. Stark, E. Schneider-Scherzer, and H. J. Thiel.** 1993. Identification of a structural glycoprotein of an RNA virus as a ribonuclease. Science **261:**1169–1171.
24. **Stark, G. R.** 1965. Reactions of cyanate with functional groups of proteins. 3. Reactions with amino and carboxyl groups. Biochemistry **4:**1030–1036.
25. **Tanaka, N., J. Arai, N. Inokuchi, T. Koyama, K. Ohgi, M. Irie, and K. T. Nakamura.** 2000. Crystal structure of a plant ribonuclease, RNase LE. J. Mol. Biol. **298:**859–873.
26. **Thiel, H. J., R. Stark, E. Weiland, T. Rumenapf, and G. Meyers.** 1991. Hog cholera virus: molecular composition of virions from a pestivirus. J. Virol. **65:**4705–4712. (Erratum, **66:**612, 1992.)
27. **Trubia, M., L. Sessa, and R. Taramelli.** 1997. Mammalian Rh/T2/S-glycoprotein ribonuclease family genes: cloning of a human member located in a region of chromosome 6 (6q27) frequently deleted in human malignancies. Genomics **42:**342–344.
28. **Vliegenthart, J. F., and F. Casset.** 1998. Novel forms of protein glycosylation. Curr. Opin. Struct. Biol. **8:**565–571.
29. **Wang, X., M. Connor, R. Smith, M. W. Maciejewski, M. E. Howden, G. M. Nicholson, M. J. Christie, and G. F. King.** 2000. Discovery and characterization of a family of insecticidal neurotoxins with a rare vicinal disulfide bridge. Nat. Struct. Biol. **7:**505–513.
30. **Wengler, G., D. W. Bradley, M. S. Collett, F. X. Heinz, R. W. Schlesinger, and J. H. Strauss (ed.).** 1995. *Flaviviridae.* Springer Verlag KG, Vienna, Austria.
31. **Yang, X., and K. Moffat.** 1996. Insights into specificity of cleavage and mechanism of cell entry from the crystal structure of the highly specific *Aspergillus ribotoxin,* restrictocin. Structure **4:**837–852.
32. **Youle, R. J., and G. D'Alessio (ed.).** 1997. Antitumor RNases. Academic Press, New York, N.Y.