Hantavirus Gc glycoprotein: evidence for a class II fusion protein

Nicole D. Tischler,[1,2] Angel Gonzalez,[3] Tomas Perez-Acle,[3]
Mario Rosemblatt[1,2,4] and Pablo D. T. Valenzuela[1,2,3,4]

Fundación Ciencia para la Vida[1], Instituto Milenio MIFAB[2], Centro de Genómica y Bioinformática, Pontificia Universidad Católica[3], and Universidad Andrés Bello[4], Zañartu 1482, Santiago, Chile

Correspondence
Nicole D. Tischler
nicole.tischler@bionova.cl *or*
nicoletis@yahoo.com

Hantavirus cell entry is promoted by its envelope glycoproteins, Gn and Gc, through cell attachment and by fusion between viral and endosomal membranes at low pH. However, the role of Gn and Gc in receptor binding and cell fusion has not yet been defined. In this work, a sequence presenting characteristics similar to those of class II fusion peptides (FPs) of alphavirus E1 and flavivirus E proteins is identified within the hantavirus Gc glycoprotein. A three-dimensional comparative molecular model based on crystallographic data of tick-borne encephalitis virus E protein is proposed for the Andes virus (ANDV) Gc ectodomain, which supports a feasible class II fusion-protein fold. *In vitro* experimental evidence is provided for the binding activity of the ANDV FP candidate to artificial membranes, as demonstrated by fluorescence anisotropy assays. Taken together, these results support the hypothesis that the Gc glycoprotein of hantaviruses and of other members of the family *Bunyaviridae* directs the viral fusion activity and that it may be classified as a class II viral fusion protein.

INTRODUCTION

Hantaviruses belong to the family *Bunyaviridae*, which encompasses five genera. Four of these include 'emerging viruses', which trigger severe diseases in humans worldwide. The best known are *California encephalitis virus* (genus *Orthobunyavirus*), *Crimean-Congo hemorrhagic fever virus* (genus *Nairovirus*), *Hantaan virus* (genus *Hantavirus*) and sandfly fever viruses (genus *Phlebovirus*). The genus *Tospovirus* is the only plant-associated genus in this family. With the exception of hantaviruses, which are transmitted by persistently infected rodents, bunyaviruses are transmitted by arthropods (Elliott, 1990).

A common feature of the family *Bunyaviridae* is the presence of two glycoproteins, which are anchored in the viral envelope membrane by their C-terminal transmembrane regions. These envelope proteins are derived from a single open reading frame of the genomic single-stranded (−) RNA medium-size segment and their size and location in the open reading frame varies between and within genera in the family *Bunyaviridae* (Elliott, 1990). After translation, the

glycoprotein precursor (GPC) is cleaved into two glycoproteins, termed Gn and Gc according to their position in the precursor. It has been suggested that these glycoproteins associate as heterodimers and accumulate in the Golgi apparatus, where viral budding takes place (Antic *et al.*, 1992; Chen & Compans, 1991; Persson & Pettersson, 1991).

Viral glycoproteins anchored to the envelope membrane are responsible for receptor recognition and entry into target cells through fusion between viral and cellular membranes. After binding to a receptor (Gavrilovskaya *et al.*, 1998; Kim *et al.*, 2002), hantaviruses enter cells through clathrin-dependent receptor-mediated endocytosis (Jin *et al.*, 2002) and are thought to fuse with endosomal membranes when the pH is below 6·3 (Arikawa *et al.*, 1985; McCaughey *et al.*, 1999). Although the fusogenic activity of hantavirus glycoproteins has been demonstrated, its assignment to Gn or to Gc has not been resolved (Ogino *et al.*, 2004).

It has been proposed that the active centre of viral fusogenic proteins consists of fusion peptides (FPs), which drive the initial partitioning of the fusion protein into the target membrane and subsequently disrupt the bilayer architecture (reviewed by Epand, 2003; Nieva & Agirre, 2003). Based on high-resolution X-ray diffraction data, two fusion machineries have recently been identified (Jardetzky & Lamb, 2004). Class I encompasses fusion proteins of several unrelated viral families, among them the influenza virus haemagglutinin, the human immunodeficiency virus gp41,

the paramyxovirus F and the Ebola virus GP2. Their common structural characteristics include a trimeric coiled-coil fold adjacent to the N-terminally located fusogenic unit, which is composed of amino acids in α-helical conformation (reviewed by Skehel & Wiley, 2000). In contrast, class II fusion proteins are distinguished by three domains of antiparallel β-sheet structures (Rey et al., 1995), containing an internal FP, which is formed by a loop flanked by two β-sheets (Allison et al., 2001; Levy-Mintz & Kielian, 1991). This second fusion class has been described for alphavirus E1 and flavivirus E proteins, belonging to the families Togaviridae and Flaviviridae, respectively. In spite of the differences between these fusion classes, FPs share several common physico-chemical and topological parameters (reviewed by Epand, 2003; Hernandez et al., 1996; Nieva & Agirre, 2003; White et al., 1983), such as high sequence conservation within the viral family, a length of 15–25 residues, a high Gly residue content and location in the ectodomain of envelope proteins.

Here, we identify and characterize an FP candidate sequence within the Gc envelope glycoprotein of hantaviruses and of associated genera. In addition, a three-dimensional molecular-model structure derived for the Gc ectodomain supports the compatibility of the hantavirus Gc glycoprotein with a class II fusion-protein fold. These results suggest a role of Gc in fusion and associates hantaviruses with the class II viral fusion machinery.

METHODS

Sequence patterns and structural annotations. To achieve a complete annotation for the GPC of the human Chilean isolate CHI-7913 of the hantavirus variant *Andes virus* (ANDV) (Tischler et al., 2003; GI, 30313865), the whole sequence was scanned for transmembrane helices by using the TMHMM version 2.0 program (Möller et al., 2001). Secondary structures were predicted as consensus from nine different algorithms by using NPS@ (Combet et al., 2000). Glycosylation sites were predicted by using the PROSITE database (Gattiker et al., 2002). To identify conserved sequence regions among members of the family *Bunyaviridae*, the block-search method was employed by using the Block Maker server (http://blocks.fhcrc.org/). The MOTIF (Smith et al., 1990) and Gibbs sampler (Lawrence et al., 1993) algorithms were used independently for the block search. Therefore, GPCs of the five genera in the family *Bunyaviridae* were extracted from the NCBI protein-sequence database. Repeated sequences (100 % similarity values) were discarded. The sequence set selected for the search comprised 30 entries (six sequences per genus). In order to assure variability, the most divergent sequences in terms of identity were used. The following accession numbers refer to the NCBI protein-sequence database: *Hantavirus*: 138339, 138342, 38505526, 38505532, 442409, 30313865; *Nairovirus*: 21326947, 14029592, 37730126, 401357, 59380666, 41052467; *Orthobunyavirus*: 138335, 14602468, 39577664, 5823127, 30409710, 30409712; *Phlebovirus*: 75198, 138341, 138345, 973315, 1174956, 52673232; *Tospovirus*: 51848026, 57157142, 18157545, 20564190, 2522489, 465409. Multiple sequence alignments were performed within each genus (six sequences), as well as with all 30 sequences of members of the family *Bunyaviridae*, by using the Blosum62 scoring matrix as implemented in CLUSTAL W (Thompson et al., 1994).

Hidden Markov models (HMMs) of hantavirus Gc, alphavirus E1 and flavivirus E proteins were extracted directly from the PFAM database seed groups (Bateman et al., 2004). HMM graphical representation was performed by HMM Logos (http://logos.molgen.mpg.de/), which incorporates both emission and transition probabilities in a graphical manner (Schuster-Böckler et al., 2004).

Fold-recognition methods. The ANDV GPC sequence was divided into the glycoproteins Gn and Gc, according to the 'WAASA' cleavage site (Löber et al., 2001). From the resultant Gc protein (487 aa), transmembrane and adjoining regions were excluded to reduce the false-positive ratio of fold-recognition programs and, hence, the 414 N-terminal residues were submitted to the following fold-recognition programs: 3D-PSSM version 2.6.0 (Kelley et al., 2000), LOOPP version 3.00 (Meller & Elber, 2001) and FUGUE version 2.0 (Shi et al., 2001). Outputs were ranked according to each program's score values and the results were subsequently inspected for cross-matches.

Comparative molecular modelling. Hydropathicity-profile comparisons were conducted through the protein hydrophilicity/hydrophobicity comparison server (http://bioinformatics.weizmann.ac.il/hydroph/). Hydropathicity indices were obtained according to the Kyte–Doolittle scale (Kyte & Doolittle, 1982) using a 21-residue window size and allowing gap insertion to encourage divergent comparisons.

MODELLER-6 (Šali & Blundell, 1993) was used to develop several comparative models of the ANDV Gc ectodomain (residues 1–414), using as template the E protein crystallographic data (Rey et al., 1995) of *Tick-borne encephalitis virus* (TBEV) (PDBid: 1SVB). The input local alignment was optimized manually to maximize the secondary-structure overlap (15·3 % identity; 24·4 % similarity). Alternatives for disulfide bridges of the model were defined heuristically by HMM comparisons of Gc and class II fusion proteins and the alignment proximity of Cys residues with the template. Twenty models were generated and ranked according to analysis of their stereochemistry by using PROCHECK (Laskowski et al., 1993). Each model was additionally ranked by the VERIFY3D score (Eisenberg et al., 1997) and potential energy values were computed by MODELLER. From the ensemble, the top 10-ranked structures were selected as starting structures for molecular-dynamic simulations.

Molecular-dynamic simulations. Template and model structures were subjected to molecular-dynamic simulations by using the GROMOS-96 force field (Van Gunsteren et al., 1996) within the Gromacs 3.1 software (Van der Spoel et al., 2002). Structures were embedded in a solvent box with the simple-point charge water model to obtain a periodic boundary condition. Long-range electrostatics were calculated with the particle-mesh Ewald method. Lennard-Jones and short-range Coulombic interactions were both cut off at 0·9 nm. Simulations were performed under normal pressure and temperature conditions, applying a constant pressure of 1 bar independently in all three directions with a coupling constant of 0·5 ps and compressibility of $4·5 \times 10^{-5}$ bar^{-1}. Additionally, ions were added to compensate the net charge of the whole system. Temperature was controlled by independently coupling the protein, solvent and counterions in a bath at a temperature of 300 K with a coupling constant of 0·1 ps. Energy minimization to reduce close contacts was achieved through the steepest-descents minimization protocol, until the maximum force decayed to 100 kJ mol^{-1} nm^{-1}. The system was then equilibrated to 300 K over 100 ps with 2 fs integration steps, using the LINCS algorithm to restrain all bond lengths (Hess et al., 1997). After this, constraints over bonds were removed and a full 1 ns molecular-dynamic simulation at 300 K was performed with 1 fs integration steps. Trajectory frames were collected each 1 ps and root mean square deviation of backbone atoms and root mean square fluctuation per residue were calculated.

Liposome preparation. Liposomes were prepared following the method of Hope *et al.* (1985). Briefly, dried lipid films were hydrated with 5 mM HEPES, 150 mM NaCl, 0·1 mM EDTA (pH 7·4) and subjected to five cycles of freezing and thawing. Subsequently, vesicles were sized by extrusion through a polycarbonate filter with a pore size of 0·1 μm. Liposomes consisted of a 1:1:1:1·5 molar ratio of phosphatidylcholine (from egg yolk), phosphatidylethanolamine (prepared from egg phosphatidylcholine by transphosphatidylation), sphingomyelin (from bovine brain) and cholesterol. All lipids were purchased from Avanti Polar Lipids. The concentration of liposome suspensions was determined by phosphate analysis (Böttcher *et al.*, 1961).

Fluorescence anisotropy of peptides. Peptides that represent the identified conserved fusion cd loop-region residues 103–130 of the ANDV Gc sequence were synthesized in two sizes. Sequences are as follows: Gc-cd1 (residues 104–125), FFEKDYQYETGWGCNPGD-CPGV; Gc-cd2 (residues 103–133), CFFEKDYQYETGWGCNPGDC PGVGTGCTACG. A negative-control peptide derived from the ANDV Gn protein (residues 80–94), VEWRKKSDTTDTTNA, was also used. All fluorescence measurements were performed with a Perkin-Elmer LS50 spectrofluorimeter equipped with polarizers in excitation and emission beams. Temperature was maintained at 20 °C. Small aliquots of liposomes (2·5 mM) were added to a 10 μM peptide solution. The suspension was incubated for 10 min before recording anisotropy through excitation at 295 nm (5 nm bandpass) and emission in L-format at 355 nm (10 nm bandpass). Anisotropy values were averaged from 10–15 measurements. Light scattering produced by liposomes was measured by incubation of the corresponding liposome concentrations with tryptophan and subtracted. The concentration of liposomes at half saturation was calculated by the dissociation constant K_D of the best-fitted hyperbola: $r=(r_{max} \times L)/(K_D+L)$, where r is the anisotropy of peptides at a given liposome concentration (L) and r_{max} is the anisotropy of bound peptides at liposome saturation.

RESULTS

Detection of an FP candidate in the hantavirus Gc glycoprotein

The sequence of the ANDV GPC was scanned in order to find properties that could relate it to either class I or class II viral fusion proteins. Fig. 1(a) shows a complete sequence pattern and structural annotation for the ANDV GPC, including the location of the cleavage site between the Gn and Gc glycoproteins (Löber *et al.*, 2001), predictions for transmembrane regions, ecto- and endodomains, hydrophobic amino acid clusters, glycosylation sites and consensus secondary structures. As shown, Gn exhibits a mixture of α-helical (21 %, red boxes) and β-strand (24 %, blue boxes) secondary structure. A higher content of residues in α-helical conformation is present at the C terminus, between the predicted transmembrane regions. In contrast, Gc displays a predominantly random-coil (62 %) and β-strand (23 %) arrangement, with a low content of α-helices (8 %).

To identify conserved regions within hantavirus glycoproteins, the Block database was analysed (Henikoff & Henikoff, 1994). The most representative conserved regions over a set of 75 hantavirus sequences are shown in Fig. 1(a) (block letters A–W). A broad conservation pattern is observed over the entire Gc sequences (block letters
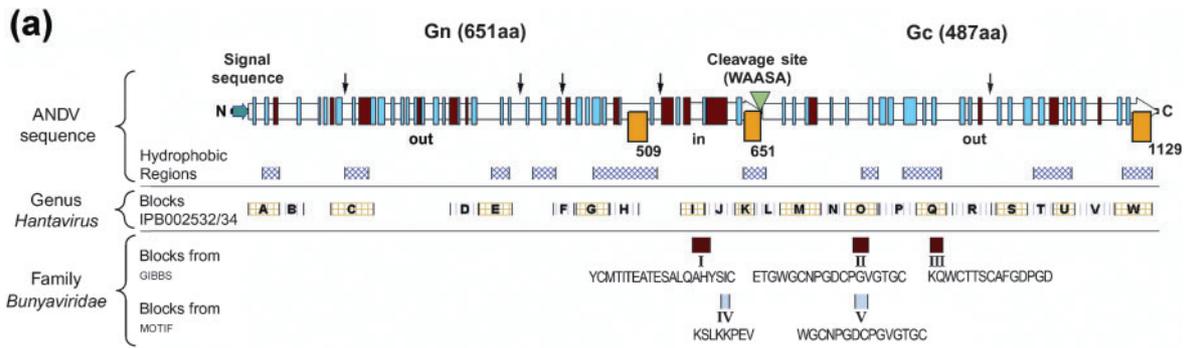
L–W), whilst a higher divergence is observed among Gn sequences (block letters A–K). Additionally, the glycoprotein sequence conservation among members of the family *Bunyaviridae*, including representatives of all five genera, was evaluated through sequence comparisons by using the Block Maker program, scanning 30 GPC sequences with two different algorithms (Gibbs and MOTIF; see Methods). Five blocks were obtained that represent regions conserved among the family *Bunyaviridae*, comprising 23 of the 30 screened GPC sequences (block numbers I–V; Fig. 1a). Six of the seven missing sequences belong to the genus *Phlebovirus*, thus eliminating this entire genus from further comparison analysis. The other missing GPC sequence corresponds to the more divergent *Dugbe virus* (NCBI protein-sequence accession no. 401357) of the genus *Nairovirus*. Nevertheless, five of the six inspected nairovirus GPC sequences included the conserved regions. The exclusion of these sequences may be explained by the use of the stringent gap penalty employed by the block algorithms.

As FPs are assumed to be conserved sequences within viral families, the five blocks conserved in the family *Bunyaviridae* were analysed for possible FP characteristics (Fig. 1a). The first (I) and fourth (IV) blocks (ANDV GPC residues 575–594 and 606–613, respectively) were discarded as FP candidates, because of their location in the predicted endodomain of Gn. The sequences of blocks II, III and V (ANDV GPC residues 763–780, 863–877 and 766–780, respectively) fulfil the requisites of conservation, ectodomain localization, minimum length and lack of putative glycosylation sites and, hence, represent suitable FP candidates.

For further studies, the sequence corresponding to blocks II and V (ANDV GPC residues 763–780) was selected based on the convergence criteria between the Gibbs and MOTIF block-search algorithms (Fig. 1a, red-underlined region in Fig. 1b). This region has 26·7 % identity and 86·7 % similarity within the analysed sequences of members of the family *Bunyaviridae*, showing conserved Cys, Gly and Trp residues (Fig. 1b). As seen in Fig. 1(a), the predicted predominant secondary structure of the FP candidate region is randomly coiled, in accordance with the high content of secondary-structure breakers such as Gly and Pro residues (eight of 15 residues within ANDV). The location of this FP candidate within the first 130 residues of Gc (ANDV Gc residues 115–129), together with the predominant β-sheet secondary structure predicted for Gc, suggests that this sequence has characteristics similar to those described for class II FPs.
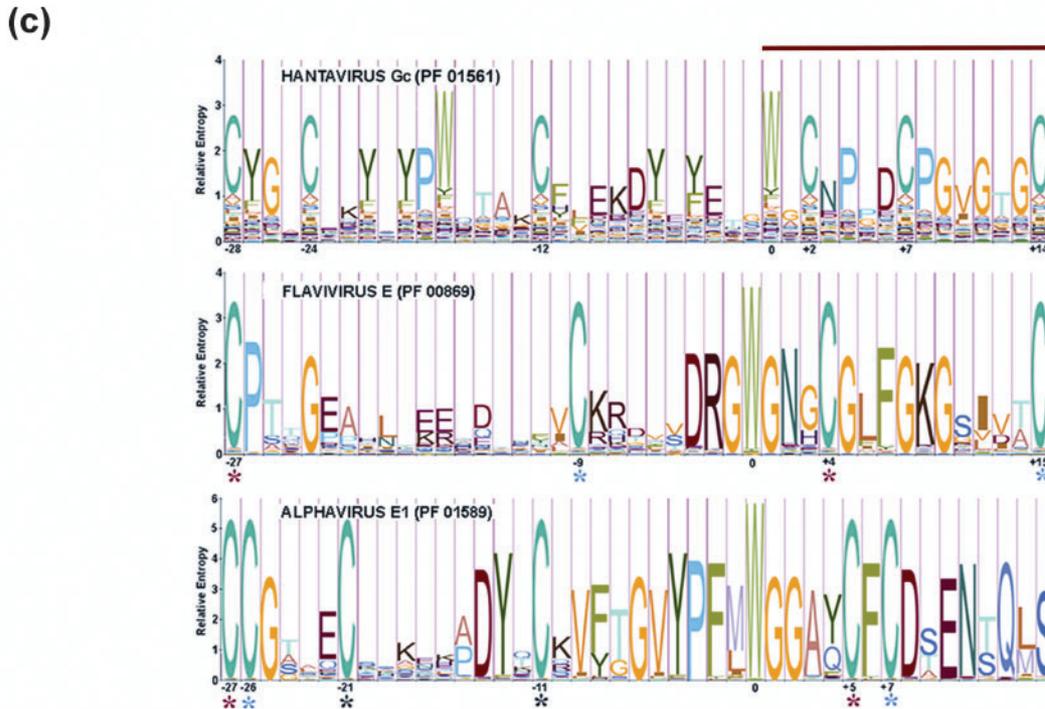
HMM comparisons of FPs

To provide further evidence that the identified sequence shares properties with class II FPs, and to obtain position-specific probabilities of the presence of key amino acids within the selected conserved region, HMMs were compared qualitatively. Fig. 1(c) (top) shows the HMM

(a)

(b)

(c)

corresponding to the hantavirus Gc region that comprises the putative FP (upper red line). The strongest position-specific probabilities belong to Trp, Cys, Gly and Pro residues. In comparison, class II FPs of flavivirus E proteins (middle) and alphavirus E1 proteins (bottom) show residues that are essential for FP functionality (see Discussion). Similarities among these three HMMs extend far beyond the putative FP region, which includes highly conserved Cys residues of alphaviruses, flaviviruses (indicated by asterisks) and hantaviruses (positions -12, -24 and -28, using the aromatic Trp 115 as reference).

Fold recognition of the ANDV Gc protein

To determine whether the ANDV Gc protein, which includes the putative FP, may adopt class II fusion-protein structural features, the fold of the first 414 residues of the ANDV Gc protein sequence was studied by using three fold-recognition programs (3D-PSSM, LOOPP and FUGUE). Results obtained with 3D-PSSM showed, as a first hit, the class II fusion protein E1 of *Sindbis virus* (PDBid: 1LD4) with an E value of 0·03 (95 % confidence). In the case of LOOPP, the dengue 2 virus fusion protein E was obtained as a hit (PDBid: 1OK8), with a threading-energy value of $-164·3$ (95 % confidence). Similar results were found with FUGUE: the Semliki Forest virus fusion protein E1 (PDBid: 1RER) had a hit with a Z-score value of 2·6 (see Supplementary Tables S1–S3, available in JGV Online). Hence, class II fusion proteins are the only cross-matched results. Furthermore, the majority of the threading output protein hits belong to the 'mainly β' class, according to the CATH database classification.

Development and evaluation of an ANDV Gc molecular model

To further analyse whether the ANDV Gc protein may support a class II fusion-protein fold, a three-dimensional comparative model was derived. In order to develop such a model, the crystallographic structure of a well-known class II fusion protein was used. To select such a structure, hydropathicity profiles among ANDV Gc and several class II fusion proteins with available crystallographic data were compared. These proteins included dengue virus E, TBEV E, Semliki Forest virus E1 and Sindbis virus E1. Despite the low sequence conservation among hantavirus Gc proteins and class II fusion proteins (approx. 20 %), an unexpectedly high hydropathic profile correspondence was found. The best matching profile for ANDV Gc was obtained with the E protein of TBEV (PDBid: 1SVB) (Fig. 2). As expected, the Gc FP candidate and the well-characterized FP of the TBEV E protein present an amphipathic nature (Fig. 2), in accordance with their requirement for partition from aqueous milieux into membranes (Nieva & Agirre, 2003). Taking into account the fact that E and E1 fusion proteins share the same topology (Lescar *et al.*, 2001), further criteria to choose the TBEV E protein as a crystallographic template were based on: (i) high-quality resolution for the crystallographic structure of the TBEV E protein at 1·90 Å; (ii) similar hydropathicity profiles along the complete sequences (Fig. 2); and (iii) consistent overlap between the predicted and observed secondary-structure elements among ANDV Gc and TBEV E proteins (Supplementary Fig. S1, available in JGV Online). Fig. 3(a) (top) shows the best-scored model of the ANDV Gc ectodomain (residues 1–414) in comparison with the TBEV E template crystallographic structure (bottom). As seen, the modelled structure retains a high content of β-sheet secondary structure along its three domains (Fig. 3a), in accordance with the two-dimensional prediction (Fig. 1a). Moreover, the FP candidate contained in the cd loop is located in an equivalent position to the FP of TBEV E, exposing the conserved tryptophan side chain. Putative disulfide bonds involve Cys residues 87–122 and 103–129, in a manner that resembles the 1SBV structure in the FP region (Fig. 3b and compare with HMMs, Fig. 1c). A potential third disulfide bond was assigned to the third domain between Cys residues 321–351, based on the

Fig. 1. Structural-domain and conserved-region analysis of glycoprotein sequences of members of the family *Bunyaviridae* and comparison with class II fusion proteins. (a) One-dimensional representation of the ANDV GPC. The WAASA cleavage site appears as a green arrowhead. Secondary-structure predictions appear as crimson red boxes for α-helices, blue boxes for β-sheets and unboxed white regions for random coils. Predicted transmembrane regions are shown in yellow boxes with a residue number indicating the terminus. Black arrows point to predicted glycosylation sites. In/out nomenclature defines the protein's endo- and ectodomains. Blue-hatched squares represent predicted hydrophobic areas. Conserved blocks within hantaviruses appear as squares containing capital letters A–W. Conserved blocks within the family *Bunyaviridae* appear in filled squares indicated by Roman numerals I–V. Their corresponding sequence in the ANDV GPC sequence is indicated below. (b) Multiple sequence alignment of the FP candidates from GPC sequences of the genera *Hantavirus*, *Orthobunyavirus*, *Nairovirus* and *Tospovirus*. NCBI protein-sequence database accession numbers appear at the beginning of each sequence, followed by the corresponding genus. Colours indicate the following: identical residues in red letters and yellow background; conserved residues in dark-blue letters and light-blue background; block of similar residues in black letters and green background; weakly similar residues in green letters and white background; and non-similar residues in black letters and white background. The red line at the bottom indicates the conserved region detected by the Block Maker program (*Bunyaviridae* blocks II and V). (c) HMMs of hantavirus FP candidates and class II FPs in Logos representation. PFAM accession numbers appear in parentheses. The relative residue position number is displayed on the *x* axis, using the conserved Trp residue as reference. Asterisks indicate the Cys residues involved in disulfide bridges according to crystal structures of flavivirus E (PDBid: 1SVB) and alphavirus E1 (PDBid: 1RER) proteins. Same-colour asterisks indicate the disulfide pairs.
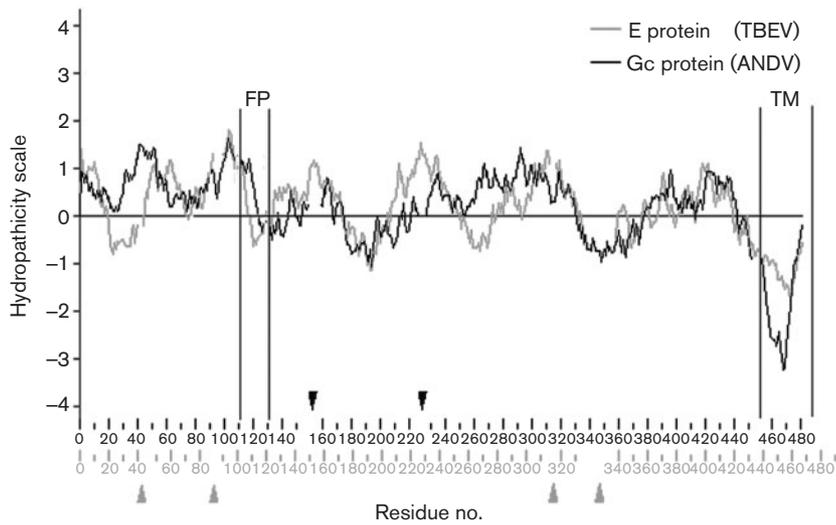
Fig. 2. Hydropathic comparison among hantavirus Gc and class II fusion proteins. Graph indicates the hydropathicity of ANDV Gc and TBEV E proteins along the entire protein length. Vertical bars indicate FP and TM regions of the E protein. Arrows indicate gap insertions in the corresponding sequences used to fit the profiles.
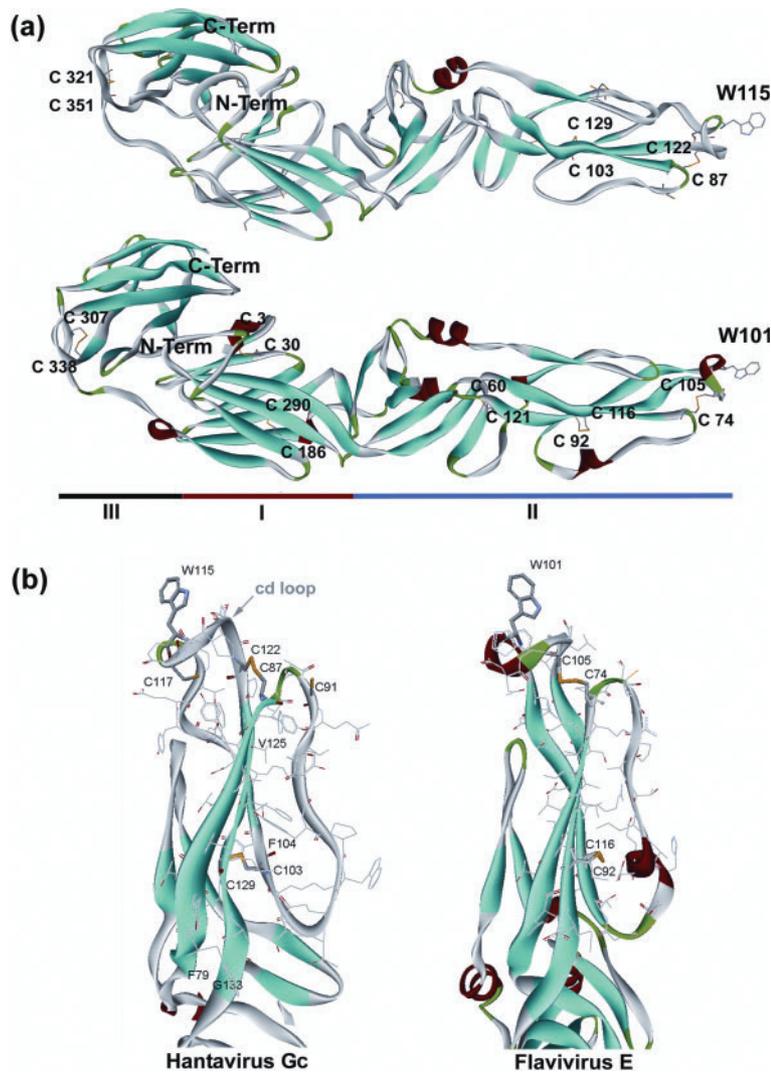
Fig. 3. ANDV Gc molecular model and comparison with template structures. (a) Backbone-structure comparison of ANDV Gc protein model (top) and TBEV E protein (1SVB) (bottom). Coloured ribbon render indicates the secondary structure using the Kabsch–Sander nomenclature (Kabsch & Sander, 1983). Cys residues involved in disulfide bridges appear yellow-labelled and stick-represented; Trp at the end of domain II in the cd loop region is also stick-represented. The horizontal bar at the bottom shows the domain organization described for class II fusion proteins. (b) Detailed view of the fusogenic domain of TBEV E and homologous domain in the ANDV Gc model.

Fig. 4. FP candidate binding to artificial membranes. Fluorescence anisotropy gain of peptides as a function of liposome concentration. Peptides (10 µM) were incubated with large unilamellar vesicles (phosphatidylcholine : phosphatidylethanolamine : sphingomyelin : cholesterol, 1 : 1 : 1 : 1·5) and anisotropy was measured at different vesicle concentrations. ●, Gc-cd1; ○, Gc-cd2; ▼, negative-control peptide; ▽, tryptophan.

proximity in the alignment of these residues with known disulfide pairs in the reference crystal structure (see Supplementary Fig. S1, available in JGV Online).
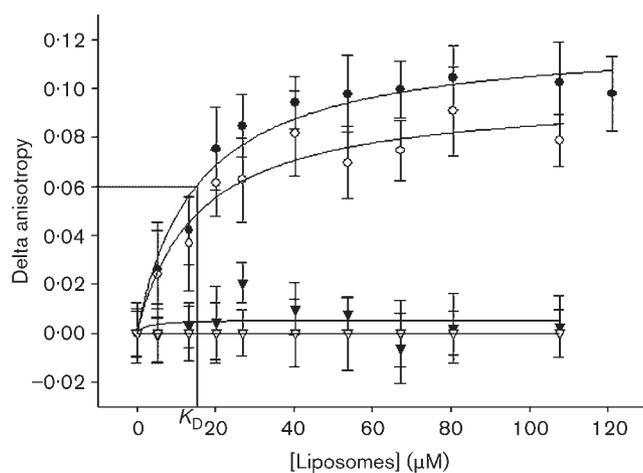
Stereochemical and dynamic methods were used to evaluate the validity of the Gc molecular model. The stereochemical quality of the Gc model was assessed by PROCHECK, which assigned 77·4 % of the residues to the Ramachandran's plot most-favoured regions, 18·9 % to additionally allowed regions, 2·5 % (nine residues) to generously allowed regions and only 1·1 % (four residues) to disallowed regions. Additional model validation using the VERIFY3D program showed an overall self-compatibility score of 114, within the expected value of 188 and the lower limit of 85 accepted for a correct fold prediction. Additional evaluations to assess model stability were performed by a 1 ns full-atom molecular-dynamic simulation for both the ANDV Gc three-dimensional model and the TBEV E reference structure (see Supplementary Fig. 2a and b, available in JGV Online).

Binding of the FP candidate to artificial membranes

To study whether the putative FP identified in the Gc protein of hantaviruses has the potential to interact with membranes, experiments with synthetic peptides in the presence of artificial membranes were performed. Given the fact that small molecules present a higher rotational movement free in solution than bound to macromolecules, the intrinsic fluorescence anisotropy of peptides was measured in their free state and in the presence of lipid vesicles. It has been shown that the fusion of flaviviruses

with artificial membranes is facilitated by the presence of sphingolipids and cholesterol (Corver et al., 2000) and that these lipids are absolutely necessary for alphavirus fusion (Waarts et al., 2002). For this reason, vesicles of two different compositions, containing only phosphatidylcholine or containing phosphatidylcholine, phosphatidylethanolamine, sphingomyelin and cholesterol, were prepared.

As seen in Fig. 3(b), the conserved FP region of class II fusion proteins (Fig. 1c) encompasses a loop, termed the cd fusion loop (Rey et al., 1995). To cover the cd fusion loop (ANDV Gc residues 103–130), two peptides of different length were synthesized (see Methods). For fluorescence experiments, the single tryptophan residue present in each peptide was employed as fluorophore. Vesicle-dependent anisotropy changes as high as 0·12 for the short Gc-cd1 peptide and 0·1 for the longer Gc-cd2 peptide were observed after incubation with vesicles made of phosphatidylcholine, phosphatidylethanolamine, sphingomyelin and cholesterol (Fig. 4). These findings show that peptides free in solution decreased their rotational movement upon addition of liposomes, reflecting their binding to these molecules. In comparison, no significant anisotropy changes were detected when a control peptide derived from the ANDV Gn sequence was incubated with the liposomes (Fig. 4).

To further support the notion that the selected peptides interact preferentially with phosphatidylcholine/phosphatidylethanolamine/sphingomyelin/cholesterol membranes, their affinities to the different liposome compositions were compared. When incubating the Gc-cd peptides with the vesicles containing the four different lipids, the K_D value derived from each curve was 15·4 µM for both Gc-cd peptides (Fig. 4). When the Gc-cd1 and Gc-cd2 peptides were incubated with vesicles containing only phosphatidylcholine, their K_D values were 65 and 96 µM, indicating a four- to sixfold lower affinity, respectively (data not shown).

DISCUSSION

Hantavirus fusion-protein identification and classification are important steps towards the development of viral cell-entry inhibitors for disease treatment. In this sense, the data presented here provide evidence for the location of an internal FP in the Gc glycoprotein of hantaviruses (Figs 1–4) and of other genera in the family *Bunyaviridae* (Fig. 1b). This peptide sequence has been described previously for its remarkable conservation within the family *Bunyaviridae* (Cortez et al., 2002; Tischler et al., 2003), but its functional relevance has not been acknowledged fully. This conserved region (ANDV Gc residues 115–129) has characteristics similar to those of known class II FPs of viruses in the families *Flaviviridae* and *Togaviridae*, in terms of conservation within the family *Bunyaviridae*, residue composition (aromatic, Gly and Cys residues), length (15–20 residues) (White et al., 1983) and secondary-structure arrangement, corresponding to a loop flanked by two β-sheets. Moreover, its location at approximately 100 residues from the N

terminus of the Gc protein, together with the predominant β-sheet secondary-structure prediction of the whole Gc protein (Fig. 1a), emphasize the class II fusion-protein features.

The presence of an internal FP in the hantavirus Gc protein is consistent with biochemical data from other members of the family *Bunyaviridae*, in which the Gc protein has been associated with the viral fusion activity. Antibodies against Gc but not against Gn of *California encephalitis virus* block syncytium formation without preventing viral attachment to the cell surface (Hacker & Hardy, 1997). In addition, an avirulent Gc variant of La Crosse virus bears a defective fusion function (González-Scarano *et al.*, 1985). Furthermore, the native Gc protein presents conformational changes at the fusion pH (Pekosz & González-Scarano, 1996), as shown for activation of other fusion proteins.

FPs have traditionally been described as hydrophobic sequences, but class II FPs also comprise charged and polar residues (see Fig. 1c). They are supposed to be anchored to the target membrane by aromatic residues and are estimated to penetrate the membrane bilayer by 6 Å (Modis *et al.*, 2004). The exposed carbonyls and charged residues on the outside rim of the fusion loop are thought to impede further penetration and may interact tightly with the phospholipid heads (Gibbons *et al.*, 2004; Modis *et al.*, 2004). This concept coincides with the observation that fusion of class II viruses with liposomes is a non-leaky process (Smit *et al.*, 2002). Therefore, a hemifusion intermediate has been proposed, in which the outer leaflets of the interacting membranes have merged, while the inner leaflets are still apart (Smit *et al.*, 2002).

To provide additional arguments for a class II FP in hantaviruses, HMMs of well-known flaviviruses and alphaviruses were compared with the HMM of hantaviruses (Fig. 1c). It is remarkable that HMMs within the known class II FPs of alphaviruses and flaviviruses are not statistically comparable (data not shown). This might not be surprising when the low sequence conservation of approximately 20 % is considered. Therefore, their classification into class II fusion proteins is based on their unexpected structural similarity, as revealed by crystallographic analysis of alphavirus E1 and flavivirus E proteins (Gibbons *et al.*, 2004; Lescar *et al.*, 2001; Modis *et al.*, 2004; Rey *et al.*, 1995). Clearly, then, three-dimensional structures of proteins may be more conserved than suggested by sequence comparison.

Hence, sequences of class II FPs seem to be quite variable among different viral families, except for particular residues involved in their functionality. These residues seem to be more conserved in terms of physico-chemical properties, rather than in terms of sequence. In this sense, the identified FP candidate of hantaviruses shows a high conservation of the required aromatic and Gly residues necessary for interaction with membranes, as well as Cys residues that are presumably involved in structure stabilization (Fig. 1b

and 1c). The location of conserved Cys residues in the hantavirus HMM in positions homologous to those of Cys residues of class II FPs far beyond the conserved FP candidate region may imply similar roles (Fig. 1c). Cys residues included in class II FPs are known to stabilize the fusion loop, which is located at the tip of domain II (Rey *et al.*, 1995; Lescar *et al.*, 2001), by three to four disulfide bridges. In summary, the similarity in sequence composition between the hantavirus FP candidate and class II FPs reinforces the hypothesis that the hantavirus Gc protein may contain a class II-like FP.

Evidence for the functionality of the identified FP candidate region is provided by its potential to interact with artificial membranes. The intrinsic fluorescence anisotropy assays clearly demonstrated the interaction of synthetic hantavirus FP candidates with lipid vesicles (Fig. 4). Their higher affinity to liposomes containing sphingomyelin and cholesterol coincides with membrane compositions known to facilitate fusion (Corver *et al.*, 2000) or to be required for the fusion of class II viruses with liposomes (Waarts *et al.*, 2002).

To study whether the hantavirus Gc protein may adopt a class II fusion-protein fold, a three-dimensional model was derived. Despite the low sequence similarity between hantavirus Gc and class II fusion proteins (approx. 20 %), encouraging results on prediction of class II fusion proteins by three different fold-recognition programs (Supplementary Tables S1–S3, available in JGV Online) persuaded model development. Although no unique set of outputs was generated by these programs, these results strongly support the notion of a β-fold structure of Gc and suggest the compatibility of the ANDV Gc sequence with a class II fusion-protein fold in terms of energy values and secondary-structure arrangement. In addition, a strong correspondence of hydropathicity profiles and secondary structures between hantavirus Gc and class II fusion proteins was observed, achieving the best consensus with the TBEV E protein (Fig. 2 and Supplementary Fig. S1, available in JGV Online).

The ANDV Gc model satisfies the acceptance requirements of applied stereochemical parameters and structural stability by means of molecular-dynamic simulations with backbone-coordinate deviations below 3 Å in a 1 ns trajectory analysis (Supplementary Fig. S2, available in JGV Online). As, in the Gc model, some Cys residues appear unpaired despite their close location in the three-dimensional space, further model improvements need to be focused on disulfide-pair formations (Supplementary Fig. S3, available in JGV Online). However, trajectory-analysis data indicate that three proposed disulfide bridges seems to be the minimal number that guarantees the structural stability of the proposed model. In summary, these results suggest that the hantavirus Gc protein is compatible with a class II fusion-protein fold and confirm its possible association with this class of viral fusion machinery.

The association of bunyavirus Gc proteins with class II fusion proteins coincides, moreover, in their overall arrangement. Class II fusion proteins are synthesized as polyproteins together with an N-terminal companion glycoprotein that acts as a chaperone to prevent their aggregation (Marquardt & Helenius, 1992) and probably suppresses their activation in the Golgi network (Guirakhoo et al., 1991, 1992), as is the case for p62 in alphaviruses (Andersson et al., 1997) and prM in flaviviruses (Konishi & Mason, 1993; Lorenz et al., 2002). Such a role could also be ascribed to bunyavirus Gn proteins, as Gc does not enter the Golgi apparatus when expressed in the absence of Gn (Persson & Pettersson, 1991; Shi & Elliott, 2002).

Class II fusion proteins are not merely responsible for fusion processes, as they also determine the viral envelope protein shell of icosahedral symmetry in either homodimeric or heterodimeric associations (Strauss & Strauss, 2001). Hence, the proposed similarity of Gc proteins of members of the family Bunyaviridae to class II fusion proteins would also influence the viral morphology. In line with this notion, regularly spaced surface projections have been described for some bunyavirions (Martin et al., 1985; McCormick et al., 1982; White et al., 1982) and even an icosahedral surface organization has been proposed for these viruses (von Bonsdorff & Pettersson, 1975; Ellis et al., 1981; Lee & Cho, 1981). Furthermore, based on the observation that the proteolytic removal of the glycoproteins produces highly deformable virion shapes, it has been hypothesized that the structural stability of bunyavirions might be conferred by the spike glycoproteins themselves (von Bonsdorff & Pettersson, 1975). In summary, and due to the absence of a matrix protein that mediates the association and stabilization between viral envelope and nucleocapsid, a highly organized structure of the surface glycoproteins of bunyaviruses may be plausible and of clear advantage for virion stability.

In conclusion, the characteristics described here for the Gc proteins of hantaviruses and other members of the family Bunyaviridae suggest their role in cell fusion and associate them with class II fusion proteins. Furthermore, these findings raise the question of whether Gn and Gc envelope glycoproteins may be involved in distinctive roles, such as receptor binding, nucleocapsid interaction and fusion, as described for alphavirus E1 and E2 proteins. To confirm the participation of the proposed FP in the hantavirus fusion activity, additional studies are required, including site-directed mutagenesis. Finally, the proposed three-dimensional Gc model may be of value in the development of cell-entry inhibitors that could be useful in therapy.

ACKNOWLEDGEMENTS

REFERENCES

Allison, S. L., Schalich, J., Stiasny, K., Mandl, C. W. & Heinz, F. X. (2001). Mutational evidence for an internal fusion peptide in flavivirus envelope protein E. J Virol **75**, 4268–4275.

Andersson, H., Barth, B.-U., Ekström, M. & Garoff, H. (1997). Oligomerization-dependent folding of the membrane fusion protein of Semliki Forest virus. J Virol **71**, 9654–9663.

Antic, D., Wright, K. E. & Kang, C. Y. (1992). Maturation of Hantaan virus glycoproteins G1 and G2. Virology **189**, 324–328.

Arikawa, J., Takashima, I. & Hashimoto, N. (1985). Cell fusion by haemorrhagic fever with renal syndrome (HFRS) viruses and its application for titration of virus infectivity and neutralizing antibody. Arch Virol **86**, 303–313.

Bateman, A., Coin, L., Durbin, R. & 10 other authors (2004). The Pfam protein families database. Nucleic Acids Res **32**, D138–D141.

Böttcher, C. J. F., Van Gent, C. M. & Pries, C. (1961). A rapid and sensitive sub-micro phosphorus determination. Anal Chim Acta **24**, 203–204.

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

Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. (2000). NPS@: network protein sequence analysis. Trends Biochem Sci **25**, 147–150.

Cortez, I., Aires, A., Pereira, A.-M., Goldbach, R., Peters, D. & Kormelink, R. (2002). Genetic organisation of Iris yellow spot virus M RNA: indications for functional homology between the $G_{(C)}$ glycoproteins of tospoviruses and animal-infecting bunyaviruses. Arch Virol **147**, 2313–2325.

Corver, J., Ortiz, A., Allison, S. L., Schalich, J., Heinz, F. X. & Wilschut, J. (2000). Membrane fusion activity of tick-borne encephalitis virus and recombinant subviral particles in a liposomal model system. Virology **269**, 37–46.

Eisenberg, D., Lüthy, R. & Bowie, J. U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol **277**, 396–404.

Elliott, R. M. (1990). Molecular biology of the Bunyaviridae. J Gen Virol **71**, 501–522.

Ellis, D. S., Southee, T., Lloyd, G., Platt, G. S., Jones, N., Stamford, S., Bowen, E. T. W. & Simpson, D. I. H. (1981). Congo/Crimean haemorrhagic fever virus from Iraq 1979: I. Morphology in BHK_{21} cells. Arch Virol **70**, 189–198.

Epand, R. M. (2003). Fusion peptides and the mechanism of viral fusion. Biochim Biophys Acta **1614**, 116–121.

Gattiker, A., Gasteiger, E. & Bairoch, A. (2002). ScanProsite: a reference implementation of a PROSITE scanning tool. Appl Bioinformatics **1**, 107–108.

Gavrilovskaya, I. N., Shepley, M., Shaw, R., Ginsberg, M. H. & Mackow, E. R. (1998). β_3 integrins mediate the cellular entry of hantaviruses that cause respiratory failure. Proc Natl Acad Sci U S A **95**, 7074–7079.

Gibbons, D. L., Vaney, M.-C., Roussel, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M. & Rey, F. A. (2004). Conformational change and protein–protein interactions of the fusion protein of Semliki Forest virus. Nature **427**, 320–325.

González-Scarano, F., Janssen, R. S., Najjar, J. A., Pobjecky, N. & Nathanson, N. (1985). An avirulent G1 glycoprotein variant of La Crosse bunyavirus with defective fusion function. J Virol **54**, 757–763.

Guirakhoo, F., Heinz, F. X., Mandl, C. W., Holzmann, H. & Kunz, C. (1991). Fusion activity of flaviviruses: comparison of mature and

immature (prM-containing) tick-borne encephalitis virions. *J Gen Virol* **72**, 1323–1329.

Guirakhoo, F., Bolin, R. A. & Roehrig, J. T. (1992). The Murray Valley encephalitis virus prM protein confers acid resistance to virus particles and alters the expression of epitopes within the R2 domain of E glycoprotein. *Virology* **191**, 921–931.

Hacker, J. K. & Hardy, J. L. (1997). Adsorptive endocytosis of California encephalitis virus into mosquito and mammalian cells: a role for G1. *Virology* **235**, 40–47.

Henikoff, S. & Henikoff, J. G. (1994). Protein family classification based on searching a database of blocks. *Genomics* **19**, 97–107.

Hernandez, L. D., Hoffman, L. R., Wolfsberg, T. G. & White, J. M. (1996). Virus-cell and cell-cell fusion. *Annu Rev Cell Dev Biol* **12**, 627–661.

Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* **18**, 1463–1472.

Hope, M. J., Bally, M. B., Webb, G. & Cullis, P. R. (1985). Production of large unilamellar vesicles by a rapid extrusion procedure. *Biochim Biophys Acta* **812**, 55–65.

Jardetzky, T. S. & Lamb, R. A. (2004). Virology: a class act. *Nature* **427**, 307–308.

Jin, M., Park, J., Lee, S. & 7 other authors (2002). Hantaan virus enters cells by clathrin-dependent receptor-mediated endocytosis. *Virology* **294**, 60–69.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.

Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. E. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* **299**, 501–522.

Kim, T.-Y., Choi, Y., Cheong, H.-S. & Choe, J. (2002). Identification of a cell surface 30 kDa protein as a candidate receptor for Hantaan virus. *J Gen Virol* **83**, 767–773.

Konishi, E. & Mason, P. W. (1993). Proper maturation of the Japanese encephalitis virus envelope glycoprotein requires cosynthesis with the premembrane protein. *J Virol* **67**, 1672–1675.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.

Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F. & Wootton, J. C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **262**, 208–214.

Lee, H. W. & Cho, H. J. (1981). Electron microscope appearance of Hantaan virus, the causative agent of Korean haemorrhagic fever. *Lancet* **i**, 1070–1072.

Lescar, J., Roussel, A., Wien, M. W., Navaza, J., Fuller, S. D., Wengler, G., Wengler, G. & Rey, F. A. (2001). The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**, 137–148.

Levy-Mintz, P. & Kielian, M. (1991). Mutagenesis of the putative fusion domain of the Semliki Forest virus spike protein. *J Virol* **65**, 4292–4300.

Löber, C., Anheier, B., Lindow, S., Klenk, H.-D. & Feldmann, H. (2001). The Hantaan virus glycoprotein precursor is cleaved at the conserved pentapeptide WAASA. *Virology* **289**, 224–229.

Lorenz, I. C., Allison, S. L., Heinz, F. X. & Helenius, A. (2002). Folding and dimerization of tick-borne encephalitis virus envelope proteins prM and E in the endoplasmic reticulum. *J Virol* **76**, 5480–5491.

Marquardt, T. & Helenius, A. (1992). Misfolding and aggregation of newly synthesized proteins in the endoplasmic reticulum. *J Cell Biol* **117**, 505–513.

Martin, M. L., Lindsey-Regnery, H., Sasso, D. R., McCormick, J. B. & Palmer, E. (1985). Distinction between Bunyaviridae genera by surface structure and comparison with Hantaan virus using negative stain electron microscopy. *Arch Virol* **86**, 17–28.

McCaughey, C., Shi, X., Elliot, R. M., Wyatt, D. E., O'Neill, H. J. & Coyle, P. V. (1999). Low pH-induced cytopathic effect – a survey of seven hantavirus strains. *J Virol Methods* **81**, 193–197.

McCormick, J. B., Sasso, D. R., Palmer, E. L. & Kiley, M. P. (1982). Morphological identification of the agent of Korean haemorrhagic fever (Hantaan virus) as a member of the Bunyaviridae. *Lancet* **i**, 765–768.

Meller, J. & Elber, R. (2001). Linear programming optimization and a double statistical filter for protein threading protocols. *Proteins* **45**, 241–261.

Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313–319.

Möller, S., Croning, M. D. R. & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653.

Nieva, J. L. & Agirre, A. (2003). Are fusion peptides a good model to study viral cell fusion? *Biochim Biophys Acta* **1614**, 104–115.

Ogino, M., Yoshimatsu, K., Ebihara, H., Araki, K., Lee, B.-H., Okumura, M. & Arikawa, J. (2004). Cell fusion activities of Hantaan virus envelope glycoproteins. *J Virol* **78**, 10776–10782.

Pekosz, A. & González-Scarano, F. (1996). The extracellular domain of La Crosse virus G1 forms oligomers and undergoes pH-dependent conformational changes. *Virology* **225**, 243–247.

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

Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375**, 291–298.

Šali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815.

Schuster-Böckler, B., Schultz, J. & Rahmann, S. (2004). HMM Logos for visualization of protein families. *BMC Bioinformatics* **5**, 7.

Shi, X. & Elliott, R. M. (2002). Golgi localization of Hantaan virus glycoproteins requires coexpression of G1 and G2. *Virology* **300**, 31–38.

Shi, J., Blundell, T. L. & Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* **310**, 243–257.

Skehel, J. J. & Wiley, D. C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* **69**, 531–569.

Smit, J. M., Li, G., Schoen, P., Corver, J., Bittman, R., Lin, K.-C. & Wilschut, J. (2002). Fusion of alphaviruses with liposomes is a non-leaky process. *FEBS Lett* **521**, 62–66.

Smith, H. O., Annau, T. M. & Chandrasegaran, S. (1990). Finding sequence motifs in groups of functionally related proteins. *Proc Natl Acad Sci U S A* **87**, 826–830.

Strauss, J. H. & Strauss, E. G. (2001). Virus evolution: how does an enveloped virus make a regular structure? *Cell* **105**, 5–8.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Tischler, N. D., Fernández, J., Müller, I. & 7 other authors (2003). Complete sequence of the genome of the human isolate of Andes virus CHI-7913: comparative sequence and protein structure analysis. *Biol Res* **36**, 201–210.

Van der Spoel, D., van Buuren, A. R., Apol, E. & 9 other authors (2002). *Gromacs User Manual*, version 3.1.1.www.gromacs.org

Van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Kruger, P., Mark, A. E., Scott, W. R. P. & Tironi, I. G. (1996). *Biomolecular Simulation: the GROMOS96 Manual and User Guide.* Zürich, Switzerland: Hochschulverlag AG der ETH Zürich.

von Bonsdorff, C. H. & Pettersson, R. (1975). Surface structure of Uukuniemi virus. *J Virol* **16**, 1296–1307.

Waarts, B.-L., Bittman, R. & Wilschut, J. (2002). Sphingolipid and cholesterol dependence of alphavirus membrane fusion. Lack of correlation with lipid raft formation in target liposomes. *J Biol Chem* **277**, 38141–38147.

White, J. D., Shirey, F. G., French, G. R., Huggins, J. W., Brand, O. M. & Lee, H. W. (1982). Hantaan virus, aetiological agent of Korean haemorrhagic fever, has bunyaviridae-like morphology. *Lancet* **i**, 768–771.

White, J., Kielian, M. & Helenius, A. (1983). Membrane fusion proteins of enveloped animal viruses. *Q Rev Biophys* **16**, 151–195.