J Mol Model (2006) 12: 921–929
DOI 10.1007/s00894-006-0101-7

ORIGINAL PAPER

**Surendra S. Negi · Andrey A. Kolokoltsov · Catherine H. Schein · Robert A. Davey · Werner Braun**

# Determining functionally important amino acid residues of the E1 protein of Venezuelan equine encephalitis virus

Received: 2 August 2005 / Accepted: 5 January 2006 / Published online: 11 April 2006
© Springer-Verlag 2006

**Abstract** A new method for predicting interacting residues in protein complexes, InterProSurf, was applied to the E1 envelope protein of Venezuelan equine encephalitis (VEEV). Monomeric and trimeric models of VEEV-E1 were constructed with our MPACK program, using the crystal structure of the E1 protein of Semliki forest virus as a template. An alignment of the E1 sequences from representative alphavirus sequences was used to determine physical chemical property motifs (likely functional areas) with our PCPMer program. Information on residue variability, propensity to be in protein interfaces, and surface exposure on the model was combined to predict surface clusters likely to interact with other viral or cellular proteins. Mutagenesis of these clusters indicated that the predictions accurately detected areas crucial for virus infection. In addition to the fusion peptide area in domain 2, at least two other surface areas play an important role in virus infection. We propose that these may be sites of interaction between the E1–E1 and E1–E2 subdomains of the envelope proteins that are required to assemble the functional unit. The InterProSurf method is, thus, an important new tool for predicting viral protein interactions. These results can aid in the design of new vaccines against alphaviruses and other viruses.

S. S. Negi · C. H. Schein · W. Braun
Sealy Center for Structural Biology,
Department of Biochemistry and Molecular Biology,
University of Texas Medical Branch,
Galveston, TX 77555-0857, USA

A. A. Kolokoltsov · R. A. Davey
Department of Microbiology and Immunology,
University of Texas Medical Branch,
Galveston, TX 77555-1075, USA

W. Braun (✉)
University of Texas Medical Branch,
2.134 Clay Hall, 0857, 301 University Boulevard,
Galveston, TX 77555, USA
e-mail: webraun@utmb.edu
Tel.: +1-409-7476810
Fax: +1-409-7476000

**Keywords** Venezuelan equine encephalitis virus (VEEV) · Alpha virus · Protein–protein interaction · Envelope glycoprotein · Functional site prediction

## Introduction

Venezuelan equine encephalitis virus (VEEV), an enveloped, positive, single-stranded RNA virus of the Togaviridae family, genus *Alphavirus*, was first recognized as an agent causing disease in animals in the 1930s. Although its primary hosts are small animals and livestock, VEEV can spread, via infected mosquitos, to humans and cause life-threatening disease characterized by fever, chills, headache, back pain, myalgias, prostration, nausea, and vomiting. Sporadic outbreaks are common, with periodic epidemics of enzootic VEEV occurring throughout North and South America [1–3]. In 1971, an outbreak originating in South America and reaching as far north as Texas resulted in tens of thousands of cases in people and the loss of more than 200,000 horses. A more recent outbreak in Columbia and Venezuela in 1995 resulted in an estimated 90,000 infected people (CDC web site, http://www.cdc.gov/ncidod/dvbid/arbor/arbdet.htm). The overall mortality rate in humans infected with enzootic strains is 0.5–1%, with up to 20% in patients who develop encephalitis. However, epizootic strains of VEEV (I-A/B and I-C) have emerged that are much more lethal, with equine mortality rates as high as 83% [3]. For these reasons and the possibility that VEEV could be weaponized, there is increased interest in developing both improved vaccines and possible inhibitors against it and related alphaviruses.

Cell entry of all alphaviruses is mediated by two envelope proteins, E1 and E2. E1 is thought to mediate fusion with the cell membrane through a “fusion peptide” that has been delineated by mutation studies. The E2 protein, which forms spikes on the viral surface, likely binds to the cellular receptor [4, 5]. Both proteins are highly conserved within the alphaviruses, with overall 50–

55% sequence identity. Cryo-EM studies of Sindbis, Semliki forest virus (SFV), and VEEV particles show that the envelope glycoproteins are arranged on the outer surface of the virus in a similar icosahedral lattice [5–8]. While there are no high-resolution crystal structures of VEEV proteins, there are crystal structures available for the E1 protein of the closely related SFV [9]. We used the known three-dimensional (3D)-structure of SFV, a bioinformatics sequence analysis, a computational method for predicting potential interacting sites and site-directed mutagenesis to determine surface regions of the E1 proteins critically involved in the cell fusion process.

Correct interactions of the E1 protein with itself [9] and other viral proteins, particularly E2, are crucial for viral assembly and presumably for successful fusion with the cell membrane after binding to the surface receptors. The high sequence identity to the SFV-E1 protein allowed us to produce a reliable model for the VEEV E1 protein, using our modeling software suite MPACK (http://curie.utmb.edu/mpack/). We then used our recently developed method for predicting interacting surfaces, InterProSurf (http://curie.utmb.edu/prosurf.html), to identify residue clusters on the surface of the model that were likely to be involved in cell-receptor and E2 interactions. Alanine substitutions were then made in the VEEV E1 protein and these were used to produce VEEV-envelope pseudotyped viruses, which bear the envelope proteins of VEEV on the core of a murine retrovirus particle [10]. Particle incorporation and infection efficiency was then measured. Our results show that mutations at residue positions predicted by InterProSurf to interact were much more likely to negatively impact viral infection than those predicted by simpler analytical methods, such as hydropathy prediction and manual surface analysis alone. The most important of these were clustered at the tip of the E1 protein and at two other positions on opposite sides of the E1 protein. The tip is most likely directly responsible for mediating membrane fusion while the other sites are more likely responsible for E1–E1 and E1–E2 interactions, but not in engaging receptor. The utility of our method in functional analysis of protein–protein interactions within the VEEV envelope proteins is discussed.

## Materials and methods

### Homology modeling and sequence analysis

We used the MPACK [11–15] suite to build a homology model of the VEEV-E1 envelope protein, using as template the crystal structure of the Semliki forest virus E1 envelope protein (PDB id 1RER, resolution 3.2A), which is ~54% identical in sequence (Figure 3 in Appendix). All disulfide-bonded cysteine residues in the SFV E1 are conserved in the VEEV E1. MPACK combines the programs EXDIS [13] to extract the distance and angle constraints from the template and DIAMOD, which generate the homology model of the protein by using the geometric constraint from EXDIS. The final model was energy minimized with FANTOM [16] and the geometry of the final model was evaluated using PROCHECK [17] (Fig. 1a). The trimeric structure of VEEV E1 was obtained by fitting homology-modeled VEEV E1 structure into the Semliki forest virus E1 trimer. The final trimer structure was energy minimized with AMBER force field. Graphics were generated with MOLMOL [18]. To model the 51 residues in the transmembrane region of E1, [19] template PDB structures 2IFO and 1IFP were selected from the fold recognition server [20]. The JPRED [21] analysis shows that the amino acid residues in the transmembrane segment are mainly hydrophobic residues and form a helix.

### Prediction of interacting sites on VEEV E1 using InterProSurf

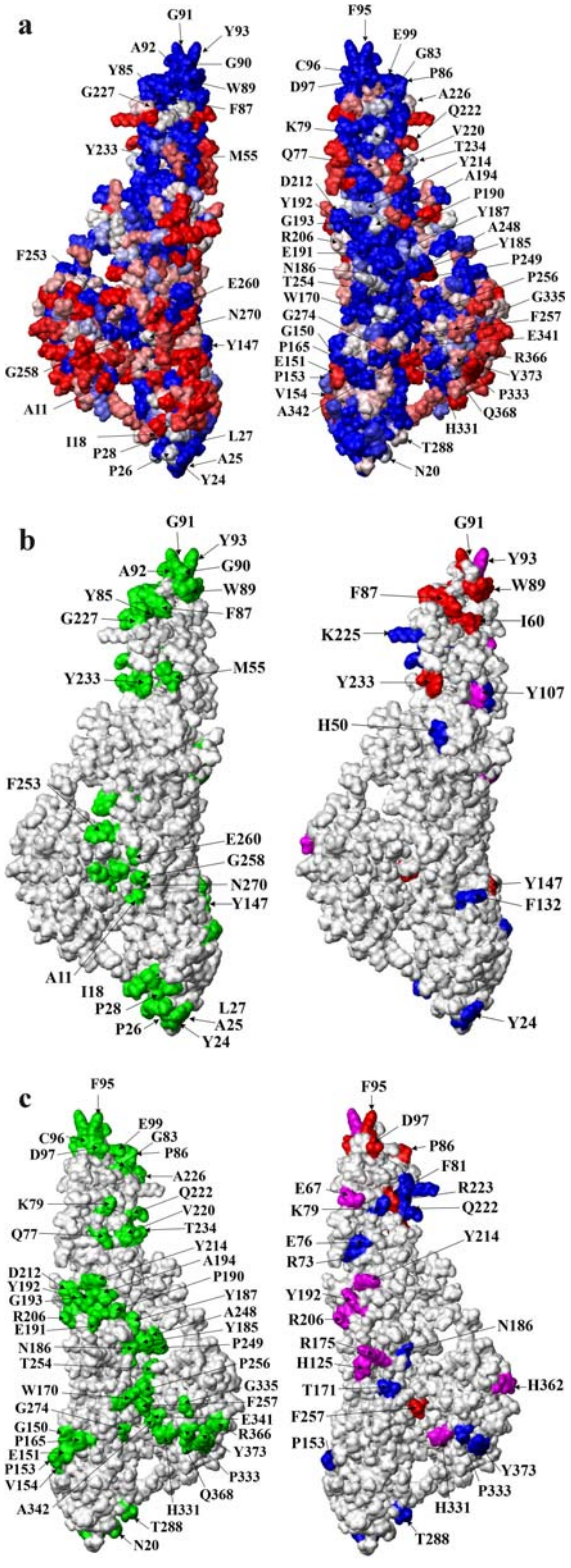
Groups of residues which are in spatial proximity on the surface of the 3D model of the VEEV E1 envelope protein were identified by using a clustering technique [22–28]. The clustering of the amino acid residues on the protein surface was based on the solvent accessible surface area of each amino acid residue calculated by the GetArea (http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl) [29]. Based on our analysis, only the amino acid residues having the side chain surface area to random coil (RSRC) value greater than 20% were assumed to be surface exposed and retained in the protein structure. The amino acid residues having RSRC values less than 20% were assumed to be buried and removed from the structure. The random coil value (RSRC) of a residue X is the average solvent accessible surface area of residue 'X' in the tripeptide Gly-X-Gly in an ensemble of 30 confirmations. In this way, all solvent-exposed residues on the protein surface were identified. In the next step, the solvent-exposed amino acid residues were replaced by their $C_\beta$ atom ($c_\alpha$ atom in case of the Gly residue). These amino acid residues on the E1 surface were clustered in such a way to minimize the distortion which is defined as square of the euclidean distance, $d(x, y)$ between the residue position ($x$) and the centroid of the cluster ($y$) [23, 28]. This can be achieved by defining an encoding region or the boundary of the cluster (e.g., $j$) as

$$V_j = \{x : d(x, y_j) < d(x, y_i) \forall i \neq j\} \quad (1)$$

The protein surface was partitioned into thirty-two clusters and the score of each cluster was calculated by

$$Score = \frac{\sum_{j \in Vj} p_j ASA_j}{\sum_{j \in V_j} ASA_j} \quad (2)$$

where $p_j$ is the propensity of amino acid residues at the protein interface and $ASA_j$ is the solvent accessible area of the amino acid residues in the unbound protein. The




◀ **Fig. 1** The homology model of VEEV E1 protein obtained by MPACK. **a** The variability plot of VEEV E1, showing the conservation of amino acid residues in their evolutions. The *blue color* indicates highly conserved residues while *red* indicates less conserved residues. The amino acid residues labeled with their *one-letter code* and *numbers* are predicted as functionally important residues. **b**, **c** Comparison of the residues predicted to be involved in protein interactions according to InterProSurf (*left*, *green* ■) with the residues effecting the titer of pseudotyped MLV particles. Color indicates the most deleterious (*red* ■) to intermediate (*magenta* ■) to wild type (*blue* ■)

clusters were sorted based on this interface score and the highest scoring clusters were predicted as being part of an interacting surface. We tested the sensitivity and precision of our prediction method for 72 test protein complexes with known 3D-structures. The sensitivity measures the ratio of all correctly predicted interfaces residues relative to all actually present interface residues, and the precision measures the ratio of all correctly identified interface residues relative to all predicted residues. If we accept only a small number of high scoring clusters in our prediction, we find a high precision and yet a low sensitivity. For eight to ten clusters, we found that our method gives a good compromise between sensitivity and accuracy. In addition to the original data set of 72 protein complexes, we also tested the performance of our method to predict the interface residue in 21 protein complexes which were not present in the training data set. The overall accuracy was found to be around 70% (Negi et al., in preparation).

## Assembly of VEEV-envelope-pseudotyped viruses and titer determination

VEEV-envelope-pseudotyped retroviruses were assembled as described previously [10]. These particles bear the envelope proteins of VEEV on the core of the retrovirus, murine leukemia virus (MLV). Virus binding to cells and infection is mediated by the VEEV envelope proteins. In brief, 293 cells were co-transfected with plasmids encoding murine leukemia virus gag and pol genes (pGAG-POL), pψ EGFP [encodes green fluorescent protein (GFP) with retrovirus packaging sequence], and pVEEV-env (encodes the envelope proteins of VEEV under control of a CMV promoter). Transfection was by calcium phosphate. Two days later, the virus was collected from culture supernatants and filtered through a 0.45-μm filter to remove cells and debris. Virus titers were determined by limiting dilution. There were 293 cells plated to 20% confluence and infected with fivefold serial dilutions of virus. Virus titer was determined by counting GFP-expressing colonies of cells. Envelope incorporation into virus particles was evaluated by Western blot analysis using the 12CA5 monoclonal antibody to detect an HA-epitope tag added to the C terminus of E1 protein as shown in Fig. 2a.

## Plasmids

All plasmids were purified by either cesium chloride density gradient or Qiagen (Valencia, CA, USA) midi columns by standard methods. The VEEV envelope expression construct is for the 3,908 subtype 1C strain of VEEV and was reported previously [10].

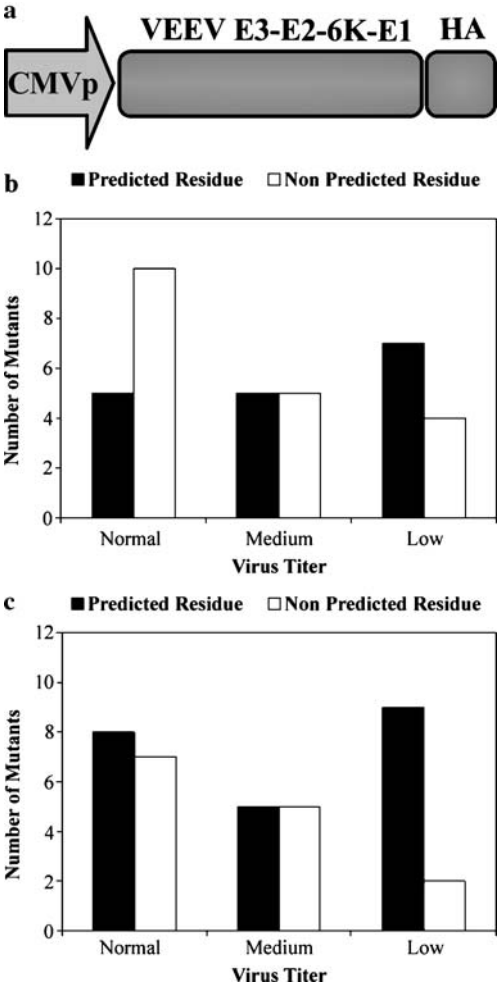



**Fig. 2** Analysis of the effect of mutations in VEEV-E1 on virus titer. Alanine was substituted for 22 surface-exposed residues predicted to be in an interface by InterProSurf (*solid bars*), as compared to 14 randomly selected amino acid residues in the E1 monomer (*open bars*). **a** Construct design (*top*). The envelope proteins of VEEV were expressed using a CMV promoter-driven expression plasmid in which E1 was modified by addition of a C-terminal HA tag. **b** Distribution of the titers for the 17 residues from the eight highest-scoring clusters and **c** for all 22 predicted residues mutated, which were selected from the ten highest-scoring clusters (*bottom*)

### Mutagenesis of E1

Amino acid substitutions were made in the E1 envelope protein using oligonucleotide-mediated site-directed mutagenesis. Codons for aromatic residues in residue clusters identified by InterProSurf were targeted and were changed to codons for alanine while residues 153, 186, 206, 233, 257, 288, 331, 214, 333, and 373 were changed to glutamine. A Quickchange kit (Stratagene) was used. All changes were confirmed by DNA sequencing.

## Results

### Description of the model

The homology model of VEEV E1, based on the crystal structure of the SFV-E1 (PDB id 1RER) [9] is shown in Fig. 1a. The root mean square deviation between the model and the template structure was 0.328 Å, consistent with the homology between the target and the template. The total area of the modeled VEEV E1 envelope protein is 20,308.71 $Å^2$ (SFV=20,475 $Å^2$), the number of surface atoms is 1,838 (SFV=1,793), and the number of buried atoms is 1,139 (SFV=1,200). The figure shows the conservation of amino acid residues in the family of nine related alphaviruses as calculated by our PCPMer program [30, 31]. The blue color indicates highly conserved residues while red indicates less conserved residues. Conserved residues of E1 in the alphavirus family map primarily on one face. This information is used in combination with predicted areas for protein interfaces by InterProSurf to find potential functional important areas of E1.

We also prepared a model of VEEV E1 as a trimer, based on that seen in crystal structures of SFV-E1 [9]. In this model, the trimer is formed by the amino acid residues forming the beta sheets in domain 1 and the amino acid residues forming the hinge region between domain 1 and domain 2. The envelope protein structure in the fusion peptide region is stabilized by two disulfide bonds, which may be necessary for correct formation of the fusion peptide and the transmembrane domain. The fusion peptide is contained in a loop between two beta strands and enters into target cells by receptor-mediated endocytosis [9, 32]. Domain 3, which has an immunoglobulin-like fold lies at the outer surface of the trimer. Most of the conserved amino acid residues found in the monomer as well as in the trimer are located in the fusion peptide loop and in the contact region between chains of the trimer. The residues in the contact regions are involved in the binding of E1 envelope protein. The amino acid residues in fusion tip of the E1 protein do not participate in the trimer contacts. However, at neutral pH, the trimer subunits may interact with each other via fusion peptide loop [9].

We used the PCPMer program [30, 31] to identify highly conserved PCP motifs in a multiple sequence alignment of E1 proteins from selected alphaviruses (Figure 4 in Appendix). A sequence analysis revealed that both envelope proteins E1 and E2 of VEEV are likely acylated and having one acylation site in each ectodomain. The VEEV E1 envelope protein has one glycosylation site at amino acid residue N-134, a NITV motif predicted by the PROSITE [33] search. The position of the glycosylation sites in E1 and E2 envelope protein of alpha viruses show that E1 positioned tangentially on the virus surface while E2 positioned radially and form spikes on the virus surface [34]. The glycosylation sites in E1 and E2 envelope protein of alpha viruses obtained from PROSITE search are shown in Table 1.

**Table 1** Glycosylation sites in E1 and E2 envelope protein of alpha viruses obtained from PROSITE search

| | |
|---|---|
| E1 VEEV 134–137 NITV | E2 VEEV 318–321 NFTV |
| E1 RRV 141–144 NQTT | E2 RRV 200–203 NCTC |
| E1 SDV 139–142 NTTS | 262–265 NVTC |
| 245–248 NNSG | E2 SDV 196–199 NITY |
| E1 SFV 141–144 NQTV | 318–321 NFTV |
| E1 EEEV 134–137 NITY | E2 SFV 200–203 NCTC |
| E1 WEEV 139–142 NTTA | 262–265 NITC |
| 245–248 NNSG | E2 EEV 315–318 NFTV |
| E1 IOV: 141–144 NITV | E2 WEEV 196–199 NVTY |
| E1 ONV: 141–144 NITV | 318–321 NFSV |
| E1 AUV: 139–142 NSTA | 405–408 NATV |
| 245–248 NNSG | E2 IOV 263–266 NTTC |
| | 345–348 NGTA |
| | E2 ONV 263–266 NTTC |
| | 345–348 NGTA |
| | E2 AUV 197–200 NVTY |
| | 319–322 NFSI |
| | 406–409 NATV |

Only one site matching the Prosite pattern (N-X-S/T) is present in the VEEV E1 and E2 envelope proteins

### Predicting interacting residues with InterProSurf

The InterProSurf method, described in "Materials and methods", was used to predict residues on the E1 surface that are most likely to interact with other proteins. The InterProSurf method uses both the surface exposure of residues on a given protein structure and our propensity scale for the amino acids to be in a protein interface (Negi et al., in preparation) to determine clusters on the surface with a high probability of being interacting sites. The accuracy of the prediction depends on the geometry of the protein surface determined by solvent accessible surface area. The prediction method was successfully tested for a large set of PDB data (Negi et al., forthcoming) and applied to predict the interacting amino acid residues in the VEEV E1 envelope protein. The results of the prediction for the VEEV E1 surface are shown in Table 2, with their interface and surface score as calculated by Eq. 2. Consistent with a functional role, the predicted amino acid residues were also highly conserved in other alpha viruses (Fig. 1a). The InterProSurf analysis was used to select ten high-scoring clusters of amino acid residues in the VEEV E1 envelope protein surface that should be important for protein–protein interaction (Fig. 1b,c). These clusters are located at the end terminal of domain 2 and at the actual interface of the envelope protein E1. Most of the clusters (1, 4, and 6 in the fusion region of domain 2; 2, 3, and 8, in the trimer interface region of the SFV crystal structure) are in regions already identified as important for interaction by previous results. Several of the clusters, especially 7 and 9, are in conserved regions of the protein that have not previously been identified as important for interaction.

### Mutagenesis of the predicted residues

To test the validity of the prediction method, substitutions were made for 22 residues predicted as important for E1 infection, as described in "Materials and methods". Mutant E1 proteins were then tested for their effects on the infectivity of pseudotyped MLV carrying the VEEV envelope proteins and compared to a set of 14 randomly chosen surface-exposed residues. The level of E1 expression for each mutant was determined in cell pellets, and incorporation into virus particles was by Western blot. We found that most recombinants were expressed and incorporated well into virions. In contrast, we obtained a wide range of viral titers from zero up to wild type. To simplify the analysis, we divided the mutant viruses into three groups according to the virus titer: normal (20–100% of wild type), intermediate deleterious (2–20%), and those that effectively precluded infectivity (<2% of wild-type plaque-forming unit). Figure 2 and Table 3 summarize the results showing that mutations at 14 positions predicted by InterProSurf, based on top ten clustering scheme, had significantly reduced viral titer (67%), while only seven of those chosen on the basis of hydropathy (46%) did. Furthermore, only two of the randomly selected mutants (13%) and nine (43%) of the InterProSurf selected residues had less than 2% of the wild type titer. These findings support the usefulness of the prediction technique in identifying regions of functional importance such as those required for protein–protein interactions. The functional relevance of each mutation for E1–E1 and E1–E2 interactions, receptor binding, and membrane fusion will be the subject of future work.

A correlation of the InterProSurf prediction with mutagenesis results showed that the amino acid substitutions at the predicted residues are more likely to have a significant impact on virus titer and the envelope protein function compared to the amino acid substitutions made for randomly chosen residues. Figure 2b,c shows the comparison of the results obtained from the InterProSurf prediction analysis and the mutagenesis experiment for the top eight and ten scoring clusters, respectively. We restricted our analysis to ten high-scoring clusters because a further increase in the cluster number will increase the sensitivity but decrease the precision of the prediction method. These results confirm the validity of the hypothesis used to predict functionally important residues on the VEEV envelope E1 protein surface as shown in Table 3.

## Discussion

### Analyzing the effects of mutations using VEEV-pseudotyped MLV

We were able to use a novel methodology to determine functional residues in the VEEV E1 envelope protein. VEEV can cause lethal human infections and there is no

**Table 2** List of amino acid residues predicted on the VEEV-E1 surface using InterProSurf

| Cluster number | Residue name | Residue number | Surface-exposed area (ASA) | Amino acid score=interface propensity×ASA | Average relative entropy | Interface score | Surface score |
|---|---|---|---|---|---|---|---|
| 32 | TRP | 89 | 184.42 | 407.18 | 2.596 | 1.63 | 0.76 |
| | GLY | 90 | 69.81 | 64.46 | 1.918 | | |
| | GLY | 91 | 48.60 | 44.87 | 1.918 | | |
| | ALA | 92 | 59.49 | 51.04 | 2.275 | | |
| | TYR | 93 | 152.66 | 272.21 | 2.000 | | |
| 20 | TYR | 185 | 56.61 | 100.94 | 2.365 | 1.49 | 0.81 |
| | ASN | 186 | 38.51 | 33.73 | 2.217 | | |
| | TYR | 187 | 73.65 | 131.33 | 1.472 | | |
| | ALA | 248 | 48.87 | 41.93 | 1.656 | | |
| | PRO | 249 | 91.90 | 95.28 | 0.538 | | |
| | PHE | 253 | 103.45 | 228.41 | 0.532 | | |
| | THR | 254 | 26.19 | 19.05 | 0.958 | | |
| 8 | ALA | 11 | 26.58 | 22.81 | 1.022 | 1.43 | 0.83 |
| | TRP | 170 | 104.42 | 230.55 | 1.823 | | |
| | PRO | 256 | 35.60 | 36.91 | 2.643 | | |
| | PHE | 257 | 82.56 | 184.32 | 1.930 | | |
| | GLY | 258 | 29.72 | 27.44 | 1.918 | | |
| | GLU | 260 | 59.03 | 44.16 | 0.926 | | |
| | ASN | 270 | 39.02 | 34.17 | 1.296 | | |
| | VAL | 273 | 39.28 | 50.33 | 1.225 | | |
| | GLU | 341 | 36.55 | 27.34 | 1.883 | | |
| | ALA | 342 | 20.40 | 17.50 | 0.876 | | |
| 30 | GLY | 83 | 26.78 | 24.73 | 1.918 | 1.37 | 0.86 |
| | TYR | 85 | 69.94 | 124.71 | 2.759 | | |
| | PRO | 86 | 123.24 | 127.78 | 2.643 | | |
| | PHE | 87 | 91.70 | 204.72 | 1.930 | | |
| | ALA | 226 | 65.62 | 56.30 | 1.229 | | |
| | GLY | 227 | 45.60 | 42.10 | 1.145 | | |
| 13 | HIS | 331 | 104.12 | 137.93 | 2.611 | 1.27 | 0.90 |
| | PRO | 333 | 80.51 | 83.47 | 1.020 | | |
| | GLY | 335 | 82.69 | 76.35 | 0.882 | | |
| | ARG | 366 | 67.44 | 75.36 | 0.749 | | |
| | GLN | 368 | 41.52 | 40.54 | 1.421 | | |
| | TYR | 373 | 121.70 | 217.00 | 0.357 | | |
| 31 | PHE | 95 | 106.77 | 238.36 | 1.930 | 1.26 | 0.90 |
| | CYS | 96 | 84.04 | 120.48 | 3.330 | | |
| | ASP | 97 | 122.17 | 87.77 | 1.937 | | |
| | GLU | 99 | 101.75 | 76.12 | 2.330 | | |
| 3 | ILE | 18 | 74.46 | 105.25 | 0.930 | 1.25 | 0.90 |
| | ASN | 20 | 81.49 | 71.37 | 1.339 | | |
| | TYR | 24 | 183.30 | 326.84 | 2.365 | | |
| | ALA | 25 | 65.31 | 56.04 | 1.324 | | |
| | PRO | 26 | 65.42 | 67.83 | 2.643 | | |
| | LEU | 27 | 63.92 | 100.58 | 1.329 | | |
| | PRO | 28 | 77.98 | 80.85 | 0.496 | | |
| | THR | 288 | 79.08 | 57.51 | 1.231 | | |

**Table 2** (continued)

| Cluster number | Residue name | Residue number | Surface-exposed area (ASA) | Amino acid score=interface propensity×ASA | Average relative entropy | Interface score | Surface score |
|---|---|---|---|---|---|---|---|
| 21 | PRO | 190 | 38.47 | 39.89 | 2.643 | 1.16 | 0.94 |
| | GLU | 191 | 82.68 | 61.85 | 1.676 | | |
| | TYR | 192 | 83.59 | 149.05 | 1.931 | | |
| | GLY | 193 | 46.64 | 43.06 | 1.918 | | |
| | ALA | 194 | 52.80 | 45.30 | 1.312 | | |
| | ASP | 212 | 59.86 | 43.00 | 1.937 | | |
| | TYR | 214 | 74.73 | 133.25 | 1.397 | | |
| | ARG | 206 | 114.31 | 127.73 | 1.218 | | |
| 26 | MET | 55 | 95.40 | 140.16 | 0.910 | 1.10 | 0.96 |
| | GLN | 77 | 61.11 | 59.66 | 1.034 | | |
| | LYS | 79 | 102.02 | 57.44 | 1.331 | | |
| | VAL | 220 | 80.46 | 103.09 | 0.686 | | |
| | GLN | 222 | 89.47 | 87.35 | 0.689 | | |
| | TYR | 233 | 78.71 | 140.35 | 2.365 | | |
| | THR | 234 | 77.68 | 56.49 | 1.332 | | |
| 5 | TYR | 147 | 63.24 | 112.76 | 1.931 | 1.10 | 0.96 |
| | GLY | 150 | 40.00 | 36.93 | 1.918 | | |
| | GLU | 151 | 98.38 | 73.60 | 0.694 | | |
| | PRO | 153 | 103.24 | 107.04 | 1.488 | | |
| | VAL | 154 | 55.42 | 71.01 | 0.792 | | |
| | PRO | 165 | 58.29 | 60.44 | 2.643 | | |

Only the top ten high-ranking clusters with their interface and surface scores are shown in the table, along with their solvent accessible surface area, average relative entropy, and their interface residue propensity and surface propensity scores. The relative entropy is calculated by the equation $R_k^i = \sum_{b=1}^{5} Q(X^b) \log_2 \left( \frac{Q(X^b)}{P(X^b)} \right)$ [31]

readily available human vaccine. Thus, work with the whole virus must be done under BSL3 conditions. The use of pseudotyped viral particles allows one to work at BSL2, which greatly facilitates analysis of large groups of

**Table 3** Effect on virus titer of mutations at residues predicted to be important for protein interactions by InterProSurf

| Amino acid residue | Total ASA | Virus titer | Predicted to be important | Residue name | Total ASA | Virus titer | Predicted to be important |
|---|---|---|---|---|---|---|---|
| Y 24 | 183.30 | Normal | Yes | Y 147 | 63.24 | Low | Yes |
| H 50 | 85.30 | Normal | No | P 153 | 103.24 | Normal | Yes |
| I 60 | 62.02 | Low | No | T 171 | 42.66 | Normal | No |
| E 67 | 128.50 | Medium | No | R 175 | 142.49 | Medium | No |
| R 73 | 61.17 | Normal | No | N 186 | 38.51 | Normal | Yes |
| E 76 | 27.20 | Normal | No | Y 192 | 83.59 | Medium | Yes |
| K 79 | 102.20 | Normal | Yes | R 206 | 114.31 | Medium | Yes |
| F 81 | 22.03 | Low | No | Y 214 | 74.73 | Medium | Yes |
| P 86 | 123.24 | Low | Yes | Q 222 | 89.47 | Normal | Yes |
| F 87 | 91.70 | Low | Yes | R 223 | 117.40 | Normal | No |
| W 89 | 184.42 | Low | Yes | K 225 | 165.09 | Normal | No |
| G 91 | 48.86 | Low | Yes | Y 233 | 78.71 | Low | Yes |
| Y 93 | 152.66 | Medium | Yes | F 257 | 82.56 | Low | Yes |
| F 95 | 106.77 | Low | Yes | T 288 | 79.08 | Normal | Yes |
| D 97 | 122.17 | Low | Yes | H 331 | 104.12 | Medium | Yes |
| Y 107 | 47.67 | Medium | No | P 333 | 80.51 | Normal | Yes |
| H 125 | 101.08 | Medium | No | H 362 | 92.97 | Medium | No |
| F 132 | 97.40 | Normal | No | Y 373 | 121.70 | Normal | Yes |

The virus titers for the mutants were classified as normal (similar to wild type), medium, or low (<2% of wild type). Amino acid residues are shown by one-letter code along with their total surface-exposed area (ASA) in the model of VEEV-E1. Most of the predicted amino acid residues have low virus titer as shown in Fig. 2

mutants, as in this study. In addition, however, the use of viral pseudotypes allows us to isolate interactions specific to the envelope proteins from other processes that the mutations might affect. Infection is a multistage process and a block at any step will result in a loss of titer. The E1 protein of VEEV must interact with itself and several other proteins, including E2 and the virus capsid, and proteins in the cell membrane at several areas of its surface during infection. Other results have also indicated that the alphavirus envelope proteins must also undergo conformational changes during the binding of the virus to the cell surface through receptor. In wild-type VEEV, the envelope–capsid interactions are also important in virus assembly but are not required for pseudotype formation and entry. The use of VEEV-pseudotyped virus particles permits us to study the envelope protein interactions alone.

### Predicting interacting residues with InterProSurf

To predict the functional sites [35, 36] on the envelope protein surface, we used our recently developed method, InterProSurf. Our method combines information from the 3D structure of the protein and the propensity of the amino acid residues to be in the protein interface to predict interacting clusters. According to statistical tests, the InterProSurf method correctly identifies the interacting sites on the protein surface important for protein–protein interaction. The accuracy of the methods depends on the properties of amino acid residues involved in the protein interaction and the geometry of the protein surface determined by solvent accessible surface area of the residues. Mutants in amino acid residues Trp89 and Gly91, which are in the highest-scoring cluster, give low virus titer. The surface exposure of the Gly90 and Gly91 residues in this cluster, which is in the fusion peptide region of VEEV E1, is higher than the nearby Gly83. This may explain why replacing the Gly91 with Asp in SFV completely blocks the fusion and infection [37] while that of Gly83 has less effect. Mutation studies in influenza hemagglutinin also show that the mutation of the Gly residue in the fusion peptide (Gly1-Ser and Gly1-Val) blocks the fusion activity [38, 39]. The amino acid residues predicted in clusters 4 and 6 were found to be highly conserved in their evolution and four of them, Pro86 and Phe87 in cluster4, and Phe95 and Asp97 in cluster 6, having low virus titer were verified experimentally. Similar effects were observed for amino acid residues Tyr147 and Phe257 which have low virus titer while the residues Tyr192, Arg206, and His331 have medium virus titer. These residues are predicted by the InterProSurf and located at the actual interface of the VEEV E1 trimer obtained by fitting VEEV-E1-modeled structure into SFV E1 trimer. The PCPMer sequence analysis also indicates high relative entropy for these residues.

The clusters we have identified as vital for VEEV E1 function in domains 2 and 3 could control the interaction of E1 with itself to form a functional homotrimer or may with E2 in the VEEV heterodimer. We will continue to test the effect of mutations in these areas in other assays to more accurately determine the role of these residues in mediating VEEV entry and fusion with the cell membrane.

## Conclusion

The prediction of the functional sites on the protein surface is a difficult and challenging problem that requires more knowledge-based approaches. Our InterProSurf analysis of the amino acid residues on the VEEV E1 envelope protein surface shows that the amino acid residues in the fusion region and in the domain 1 of the envelope protein play a significant role in the fusion and the stability of the virus itself. These results were verified experimentally using alanine mutagenesis. Our findings suggest that mutations in residues in clusters predicted to be in interface clusters had significant effects on the infectivity of pseudotyped virus particles. These results help us to determine possible interacting residues responsible for E1–E1 or E1–E2 interactions. Further experimental results will aid in refining our model of the surface features of the E1 protein that play a role in viral assembly and fusion. This information, in turn, can help guide the development of new vaccines against alpha viruses and aid in the design of inhibitors of the viral fusion process.

**Acknowledgements** We thank Dr. Numan Oezguen and Dr. Bin Zhou for fruitful discussions and help in drawing the amino acid residue variability plot. This project was supported by NIH grant No. R21 AI055746.

## Appendix

```
VEEVE=VEEV E1 Sequence
1RERA=1RER PDB Sequence

VEEVE: 1   YEHATTMPSQAGISYNTIVNRAGYAPLPISITPTKIKLIPTVNLEYVTCHYKTGMDSPAI 60
           ||| | ||   |  |     | || ||          | || |||| || |||   ||
1RERA: 1   YEHSTVMPNVVGFPYKAHIERPGYSPLTLQMQVVETSLEPTLNLEYITCEYKTVVPSPYV 60

VEEVE: 61  KCCGSQECTPTNRPDEQCKVFTGVYPFMWGGAYCFCDTENTQVSKAYVMKSDDCLADHAE 120
           ||||  ||      || |||| ||||||||||||||||| |||| | |||  || |   |||
1RERA: 61  KCCGASECSTKEKPDYQCKVYTGVYPFMWGGAYCFCDSENTQLSEAYVDRSDVCRHDHAS 120

VEEVE: 121 AYKAHTASVQAFLNITVGEHSIVTTVYVNGETPVNFNGVKLTAGPLSTAWTPFDRKIVQY 180
           ||||||||  |      |       |||||   |   |     |||| |||||| ||| |
1RERA: 121 AYKAHTASLKAKVRVMYGNVNQTVDVYVNGDHAVTIGGTQFIFGPLSSAWTPFDNKIVVY 180

VEEVE: 181 AGEIYNYDFPEYGAGQPGAFGDIQSRTVSSSDLYANTNLVLQRPKAGAIHVPYTQAPSGF 240
             |  | ||| || |||| ||||||||| | |||||| | | ||  |  |||||| ||||
1RERA: 181 KDEVFNQDFPPYGSGQPGRFGDIQSRTVESNDLYANTALKLARPSPGMVHVPYTQTPSGF 240

VEEVE: 241 EQWKKDKAPSLKFTAPFGCEIYTNPIRAENCAVGSIPLAFDIPDALFTRVSETPTLSAAE 300
           | | |   |  ||||| | ||| || |||||| ||     ||  |||  | ||
1RERA: 241 KYWLKEKGTALNTKAPFGCQIKTNPVRAMNCAVGNIPVSMNLPDSAFTRIVEAPTIIDLT 300

VEEVE: 301 CTLNECVYSSDFGGIATVKYSASKSGKCAVHVPSGTATLKEAAVELTEQGSATIHFSTAN 360
           ||   | ||||||  |  |   | | | ||  |  ||| ||       |  | |||||
1RERA: 301 CTVATCTHSSDFGGVLTLTYKTNKNGDCSVHSHSNVATLQEATAKVKTAGKVTLHFSTAS 360

VEEVE: 361 IHPEFRLQICTSYVTCKGDCHPPKDHIVTHP 391
             | |    |   ||   | |||||||
1RERA: 361 ASPSFVVSLCSARATCSASCEPPKDHIVPYA 391
```

**Fig. 3** Pairwise sequence alignment of VEEV E1 envelope protein (Genbank ID 25140297) and Semliki forest E1 envelope protein (PDB ID 1RER). Overall sequence identity between VEEV E1 and SFV E1 was found to be 54%

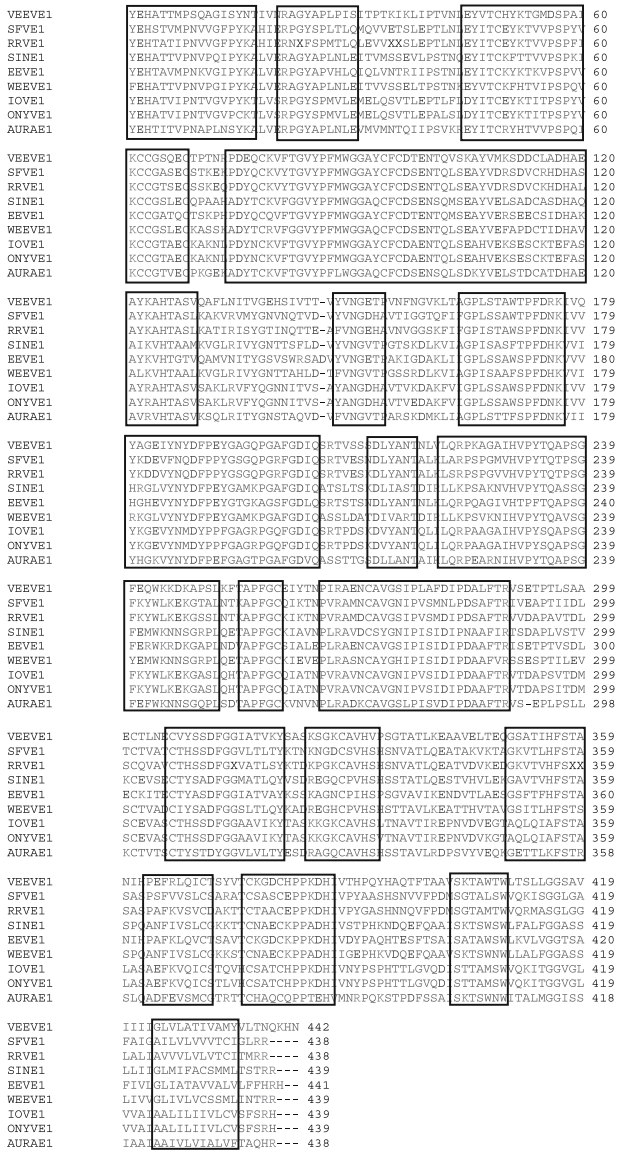


**Fig. 4** CLUSTAL W multiple sequence alignment of VEEV E1 envelope protein with other alpha family E1 protein sequences. The abbreviation used in the multiple sequence alignment are Venezuelan equine encephalitis virus (VEEVE1, Genbank no. 25140297), Semliki forest (SFVE1, Genbank no. 29612016), Ross River virus (RRVE1, Genbank no. 25121500), Sindbis (SINE1, Genbank no. 25121513), Eastern equine encephalitis (EEVE1, Genbank no. 25121478), Western equine encephalomyelitis (WEEVE1, Genbank no. 29611994), Igbo Ora (IOVE1, Genbank no. 25140289), O'nyong-nyong (ONYVE1, Genbank no. 25121491), and Aura virus (AURAE1, Genbank no. 29653361). Motif identified by PCPMer method are shown by boxes using the gap parameter, $G$=2, entropy cutoff =2, and length cutoff $L$=4

## References


1. Wenger F (1977) Teratology 16:359–362
2. Paessler S, Fayzulin RZ, Anishchenko M, Greene IP, Weaver SC, Frolov I (2003) J Virol 77:9278–9286
3. Weaver SC, Ferro C, Barrera R, Boshell J, Navarro JC (2004) Annu Rev Entomol 49:141–174
4. Kinney RM, Tsuchiya KR, Sneider JM, Trent DW (1992) Virology 191:569–580
5. Zhang W, Mukhopadhyay S, Pletnev SV, Baker TS, Kuhn RJ, Rossmann MG (2002) J Virol 76:11645–11658
6. Zhang W, Fisher BR, Olson NH, Strauss JH, Kuhn RJ, Baker TS (2002) J Virol 76:7239–7246
7. Lescar J, Roussel A, Wien MW, Navaza J, Fuller SD, Wengler G, Rey FA (2001) Cell 105:137–148
8. Paredes A, Alwell-Warda K, Weaver SC, Chiu W, Watowich SJ (2003) J Virol 77:659–664
9. Gibbons DL, Vaney MC, Roussel A, Vigouroux A, Reilly B, Lepault J, Kielian M, Rey FA (2004) Nature 427:320–325
10. Kolokoltsov AA, Davey RA (2004) J Virol 78:5124–5132
11. Soman KV, Midoro-Horiuti T, Ferreon JC, Goldblum RM, Brooks EG, Kurosky A, Braun W, Schein CH (2000) Biophys J 79:1601–1609
12. Schein CH, Nagle GT, Page JS, Sweedler JV, Xu Y, Painter SD, Braun W (2001) Biophys J 81:463–472
13. Soman KV, Schein CH, Zhu H, Braun W (2001) Methods Mol Biol 160:263–286
14. Ivanciuc O, Oezguen N, Mathura VS, Schein CH, Xu Y, Braun W (2004) Curr Med Chem 11:583–593
15. Schein CH, Zhou B, Oezguen N, Mathura VS, Braun W (2005) Proteins: Struct, Funct, Bioinf 58:200–210
16. Schaumann T, Braun W, Wuthrich K (1990) Biopolymers 29:679–694
17. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291
18. Koradi R, Billeter M, Wuthrich K (1996) J Mol Graph 14:51–55
19. Mancini EJ, Clarke M, Gowen BE, Rutten T, Fuller SD (2000) Mol Cell 5:255–266
20. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) Bioinformatics 19:1015–1018
21. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) Bioinformatics 14:892–893
22. Linde Y, Buzo A, Gray RM (1980) IEEE Trans Commun 28:84–95
23. Sayood K (2000) Introduction to data compression, 2nd edn. Kaufmann, San Francisco, CA
24. Singh RK, Tropsha A, Vaisman II (1996) J Comput Biol 3:213–221
25. Liang J, Edelsbrunner H, Woodward C (1998) Protein Sci 7:1884–1897
26. Landgraf R, Xenarios I, Eisenberg D (2001) J Mol Biol 307:1487–1502
27. De-Alarcon PA, Pascual-Montano A, Gupta A, Carazo JM (2002) Biophys J 83:619–632
28. Patane G, Russo M (2001) Neural Netw 14:1219–1237
29. Fraczkiewicz R, Braun W (1998) J Comput Chem 19:319–333
30. Schein CH, Zhou B, Braun W (2005) Virol J 2:40
31. Mathura VS, Schein CH, Braun W (2003) Bioinformatics 19:1381–1390
32. Bressanelli S, Stiasny K, Allison SL, Stura EA, Duquerroy S, Lescar J, Heinz FX, Rey FA (2004) EMBO J 23:728–738
33. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch APB (2002) Brief Bioinform 3:265–274
34. Pletnev SV, Zhang W, Mukhopadhyay S, Fisher BR, Hernandez R, Brown DT, Baker TS, Rossmann MG, Kuhn RJ (2001) Cell 105:127–136
35. DeLano WL (2002) Curr Opin Struct Biol 12:14–20
36. Bogan AA, Thorn KS (1998) J Mol Biol 280:1–9
37. Shome SG, Kielian M (2001) Virology 279:146–160
38. Han X, Bushweller JH, Cafiso DS, Tamm LK (2001) Nat Struct Biol 8:715–720
39. Tamm LK, Han X, Li YL, Lai AL (2002) Biopolymers 66:249–260