

Molecular Genetic Evidence that the Hydrophobic Anchors of Glycoproteins E2 and E1 Interact during Assembly of Alphaviruses

Ellen G. Strauss,* Edith M. Lenches, and James H. Strauss

Division of Biology, California Institute of Technology, Pasadena, California 91125

Received 23 April 2002/Accepted 8 July 2002

Chimeric alphaviruses in which the 6K and glycoprotein E1 moieties of Sindbis virus are replaced with those of Ross River virus grow very poorly, but upon passage, adapted variants arise that grow >100 times better. We have sequenced the entire domain encoding the E2, 6K, and E1 proteins of a number of these adapted variants and found that most acquired two amino acid changes, which had cumulative effects. In three independent passage series, amino acid 380 of E2, which is in the transmembrane domain, was mutated from the original isoleucine to serine in two instances and to valine once. We have now changed this residue to seven others by site-directed mutagenesis and tested the effects of these mutations on the growth of both the chimera [SIN(RRE1)] and of parental Sindbis. These results indicate that the transmembrane domains of glycoproteins E2 and E1 of alphaviruses interact in a sequence-dependent manner and that this interaction is required for efficient budding and assembly of infectious virions.

During assembly of alphaviruses in vertebrate cells, an icosahedral nucleocapsid buds through the cell plasma membrane to acquire an envelope containing two virus-encoded glycoproteins, E1 and E2 (reviewed in references 18 and 19). PE2, a precursor to E2, and E1 are synthesized on the rough endoplasmic reticulum as a precursor polyprotein and are cleaved by host cell signalase, releasing a small peptide called 6K. PE2 and E1 quickly associate to form a heterodimer that is transported through the Golgi apparatus to the cell plasma membrane. During transport, PE2 is cleaved to E2 and the hydrophobic domains of both E2 and E1 are fatty acylated. Prior to or during assembly of the virus, three E2E1 heterodimers trimerize to form a spike. Both E1 and E2 are anchored in the lipid bilayer by hydrophobic anchors near their carboxy termini. E1 has only two amino acids on the cytoplasmic side of the bilayer, whereas E2 has 33. There is genetic and structural evidence that the C-terminal cytoplasmic domain of E2 and the nucleocapsid protein interact during assembly (19). The final assembled virus has T=4 icosahedral symmetry with 240 glycoprotein heterodimers forming 80 trimeric spikes on the surface of the particle, which envelope a nucleocapsid that also has regular T=4 symmetry (1, 2, 13, 17, 19). The assembly mechanisms of alphaviruses are quite precise in comparison to those of most enveloped viruses, and the virion has been described as a regular protein lattice with a lipid bilayer embedded in it (15).

We have constructed chimeric viruses in which PE2, E1, or the nucleocapsid protein C is derived from two different alphaviruses, Ross River virus (RR) and Sindbis virus (SIN), which share only about 50% amino acid identity in their structural proteins (7, 8, 12, 22, 23). Thus, the replacement of a structural protein in one virus by one from the other virus

results in changing about half of the amino acids, but unlike what is seen in random mutagenesis, the heterologous glycoprotein is known to be functional in its native context. It was shown recently that a chimera in which the SIN E1 protein was replaced with that of RR was highly attenuated, producing virus at a rate of about 10^{-7} that of SIN. The attenuation resulted from a failure of nucleocapsids to interact with glycoprotein heterodimers in the plasma membrane (23). Continued passage of this chimeric virus led to the appearance of variants that were able to produce virus at 100-fold or better the rate of the parental chimera. Complete characterization of one such variant showed that it had acquired one mutation in SIN E2 (Asp-238→Tyr) and one in RR E1 (Gln-411→Leu), which adapted these two glycoproteins to one another (22). Ten other independent variants were partially characterized by sequencing the regions around E2 Asp-238 and E1 Gln-411. Here we report the complete sequence of the structural proteins of seven such variants and identify a number of mutations that adapt SIN E2 to RR E1. Many of these mutations were found in or near the hydrophobic domains that anchor E2 and E1 in the bilayer. In particular, several changes were found at Ile-380 of E2, which is located in the middle of the transmembrane anchor, and, to further examine the importance of this residue, nine site-specific mutations were made at Ile-380 and their effects on the growth of both the SIN/RR chimera and upon SIN itself were examined. The glycoprotein anchors of alphaviruses show only modest conservation among alphaviruses, but the results presented here indicate that E1 and E2 interact within their stem and anchor regions and that the specific sequences within these regions affect the stability of these interactions.

MATERIALS AND METHODS

Virus and cells. The characterization of the full-length cDNA clone pSIN-(RRE1) and transfection of cells with RNA from this clone have been described earlier (23). Passaging of this chimera in BHK cells to produce 11 independent passage series of variants has been described earlier (22). To determine the

* Corresponding author. Mailing address: Division of Biology, 156-29, California Institute of Technology, Pasadena, CA 91125. Phone: (626) 395-4903. Fax: (626) 449-0756. E-mail: strausse@cco.caltech.edu.

nucleotide sequence encoding the glycoproteins of the variants, RNA was isolated from purified virus and amplified by reverse transcriptase PCR and the cDNA was cloned into pGEM3Z vectors and sequenced as previously described (22).

cDNA clones with changes at Ile-380. To rule out the possible importance of changes that might have occurred elsewhere in the genome during passage of the chimeras upon the growth of the adapted chimera and to test the effects of the mutations upon the growth of SIN, the Ile-380 mutations were moved into full-length clones of SIN and of SIN(RRE1). The full-length cDNA clone of SIN, Toto 54 (13,862 nucleotides [nt] in size) was digested with *Bss*HII and *Mlu*I, and the two resulting fragments were purified from agarose gels. The 2,201-residue fragment from *Mlu*I (nt 7603) to *Bss*HII (nt 9804) was further digested with *Pst*I (nt 9115) to give a fragment of 1,512 nt (nt 7603 to 9115). The reverse transcriptase PCR subclones derived from the E2 region of adapted variants contained the C-terminal 18 amino acids of the SIN capsid protein, all of SIN E3 and SIN E2, the RR 6K protein, and the two N-terminal amino acids of RR E1. Subclones containing the mutations I380V and I380S were digested with *Pst*I and *Bss*HII, and the fragment of 689 nt containing the Ile-380 codon (nt 9766 to 9768) was joined in a three-piece ligation with the *Mlu*I-*Pst*I and *Bss*HII-*Mlu*I fragments from pToto 54 to give a full-length SIN clone.

Toto 54 constructs containing the I380V or I380S mutation were digested with *Spe*I (nt 5262) and *Bss*HII (nt 9804), and the 4,542-nt fragment was purified. Full-length SIN(RRE1) was digested with the same enzymes, and the 9,528-nt was fragment purified. These were ligated together to yield SIN(RRE1) constructs containing I380V and I380S.

Site-specific mutagenesis at Ile-380. Site-specific mutagenesis was performed according to the method of Zoller and Smith (26) as modified by Kunkel (9). The viral *Pst*I-*Pst*I fragment starting at nt 9115 was inserted into *Pst*I-digested M13mp18. The fragment from Toto 54 was 1,700 nt long, that from SIN(RR6K) was 1,715 nt, and that from SIN(RRE1) was 1,049 nt. The orientation of the inserts was checked by restriction digestion. M13 phages with the appropriate inserts from the three different backgrounds were mutagenized using a degenerate primer, 3' CGC TAC TAC NNA CCG CAT TGA 5', where N is an equal mixture of all 4 nt. M13 phages with inserts, forming clear plaques, were isolated and grown as small cultures, and the sequence at the Ile-380 site was determined on single-stranded DNA that had been isolated from phage. The double-stranded replicative form (RF) of M13 DNA was isolated from cells infected with phages containing mutations and was digested with *Pst*I and *Bss*HII. These RF-derived fragments were inserted into full-length Toto 54, SIN(RR6K), or SIN(RRE1), as appropriate, by a three-piece ligation with the *Bss*HII-*Mlu*I and *Mlu*I-*Pst*I fragments from SIN Toto 54 or the two chimeras. The full-length constructs were checked by restriction digestion, and the sequence in the region of Ile-380 was confirmed by sequencing. Mutants obtained in one construct but not in others were moved into the other constructs by swapping the *Mlu*I-*Bss*HII fragment.

Growth curves of Ile-380 mutants. Differential growth curves of the mutant constructs were performed in parallel with the parental constructs, using monolayers of BHK-21 cells. RNA was transcribed *in vitro* from the full-length cDNA clones and used to transfect BHK cells using Lipofectin or DEAE dextran, as previously described (7). At intervals, the supernatant medium was removed and replaced with fresh medium, and the harvested supernatants were titered by plaque assay on BHK monolayers.

RESULTS

Sequence of adapted variants. It was previously reported that a chimeric virus consisting of SIN in which the 6K protein and E1 were replaced by their counterparts from RR was viable but highly attenuated. PE2 and E1 formed a chimeric heterodimer that was cleaved to an E2E1 heterodimer and transported to the cell plasma membrane but failed to achieve the correct final conformation and failed to interact with nucleocapsids (23). Passage of the chimera 10 times resulted in production of variants that grew about 100-fold better than the parental chimera. One of these variants was completely characterized, and 10 other variants were partially characterized (22). We have now completely sequenced the structural protein domain of seven of the variants that were previously only partially characterized, and the results for the eight completely

sequenced variants are shown in Fig. 1. All variants have at least two changes within the glycoproteins, consistent with earlier findings that two changes were required to effect the 100-fold increase in virus replication observed with the one variant studied in detail. Six of the variants had changes in both RR E1 and in SIN E2, whereas one had changes only in SIN E2 and one had changes only in RR E1.

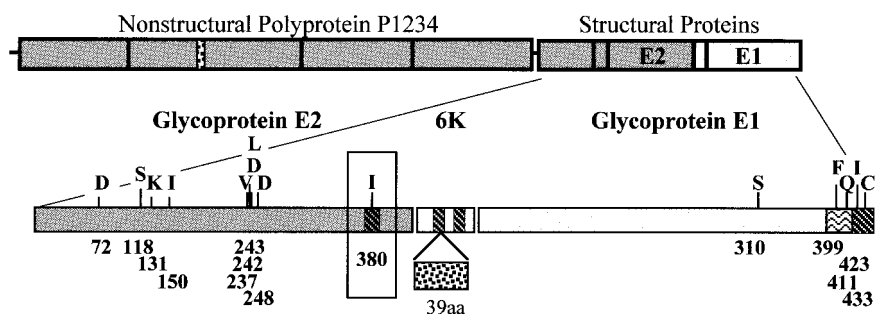
Changes in RR E1 were found in seven of the sequenced variants. The changes Ile-423→Leu and Cys-433→Arg occur within the membrane anchor of RR E1, which in RR probably spans E1 residues 413 to 435. The Cys-433→Arg change is of particular interest because it involves addition of charge and could alter the membrane-spanning domain of E1. Two changes, Gln-411→Leu, which occurs twice, and Phe-399→Ser, are found in the ectodomain of E1 immediately adjacent to the lipid bilayer, i.e., in the stem region. In the upstream ectodomain the change Phe-310→Ser occurs in three variants. Phe-310 is found in domain III of Semliki Forest virus E1 (10). The change Val-113→Ala also occurred but was not confirmed by sequencing of a second clone and thus could have arisen during PCR amplification.

The changes Q411L, S310F, and C433R were shown previously to lead to an increase in titer of the chimera and thus to be adaptive (7, 22). Because the F310S change occurred in three independent variants, it is also clearly adaptive. We assume that the change I423L is also adaptive because it appears to have been selected during passage. The importance of the V113A change is unknown, as indicated above.

Seven of the eight sequenced variants had changes in E2, and four of these had more than one change in E2. These changes occur at many places in E2 but appear to cluster into several regions, which might represent domains that interact with E1. Interestingly, three variants have changes in SIN E2 Ile-380, twice to Ser and once to Val. This change is found within the hydrophobic anchor of glycoprotein E2, which is believed to comprise residues 365 to 390. Thus, adaptive changes occurred in the membrane anchors of both E1 and E2. Other E2 changes occurred in the ectodomain. Three changes in these eight completely sequenced variants are found in the domain from 237 to 248: Val-237→Phe, Leu-243→Ser, and Asp-248→Tyr. Two other variants, which have been only partially characterized, were also found to have changes in this domain (Asp-248→Ala and Asp242→Gly) (22), and thus, five different changes in four different amino acids occurred within a stretch of only 12 residues. Two changes are found in a domain near the N terminus (Ser-60→Gly and Asp-72→Asn), and five changes are found in the domain from 108 to 150 (Gly-108→Ala, Ser-118→Asn, Lys-131→Glu, which occurred twice, and Ile-150→Leu). The changes K131E, V237F, and D248Y have been shown to be adaptive (7, 22), and as shown below, the changes at Ile-380 are also adaptive. We assume that the other changes selected during passage are also adaptive.

In addition to the changes in E2 and E1, we also identified one change in the 6K protein, an insertion of 39 amino acids between residues 42 and 43 of RR 6K. The nucleotide sequence encoding these 39 residues is derived from SIN nsP2 (as indicated in Fig. 1) and is translated in the same reading frame as in nsP2 (Fig. 2).

A. Schematic of SIN(RR6KE1) with locations of suppressor mutations



B. Changes in E2, 6K, and E1 during Passage

Passage Series	E2 changes	6K Changes	E1 Changes
1	Asp-248 → Tyr	None	Gln-411 → Leu
2	Leu-243 → Ser (Ser-60 → Gly) (Gly-108 → Ala)	39 amino acid Insertion	Ser-310 → Phe Ile-423 → Leu
3	Val-237 → Phe Ile-380 → Ser	None	None
4	None	None	Gln-411 → Leu (Val-113 → Ala)
5	Lys-131 → Glu Ile-150 → Leu Ile-380 → Ser	None	Phe-399 → Ser
8	Asp-72 → Asn Ile-380 → Val	None	Ser-310 → Phe
9	Lys-131 → Glu	None	Cys-433 → Arg
11	Ser-118 → Asn	None	Ser-310 → Phe

FIG. 1. Amino acid changes selected during passage of chimeric SIN(RRE1) virus. (A) Diagram of the genome organization of SIN(RR6KE1) with an expanded view of the glycoprotein region below. SIN sequences are shaded dark gray, and RR sequence is shaded light gray. Transmembrane domains are shaded with diagonal hatching; the "stem" region of E1 is shaded with wavy lines. Residue numbers where variants mapped are indicated below the open reading frame, and the parental amino acid in each case is shown above. aa, amino acids. (B) Amino acid changes found in eight independent passage series. In most cases the changes found were shown to be representative of the variant population by sequencing two independent clones and finding the change in each clone or by sequencing the PCR DNA prior to cloning. For the changes shown in parentheses, only a single clone was sequenced and it is uncertain whether these changes are representative of the revertant population. A limited number of other changes were also observed that were clonal, that is, found only in one of two sequenced clones and that may have arisen by PCR mutagenesis during cloning. These changes are not shown because their origin is uncertain and because, in any event, they do not represent the consensus sequence of the variant population.

Mutagenesis of Ile-380. It is striking that, in three of eight variants examined, a change occurred at E2 Ile-380. Because these changes occurred within the hydrophobic anchor of glycoprotein E2 and because two different substitutions occurred at Ile-380, we wished to examine more closely the importance of this position for the interaction of RR E1 and SIN E2. For this purpose, the serine and valine substitutions found in variants were cloned back into parental cDNA constructs to test the effects of these substitutions in the absence of other changes that might have occurred during passage. We also made and tested seven other mutations at Ile-380.

Growth of SIN(RRE1) chimeras containing each of 10 different amino acids at SIN E2 residue 380 was examined by transfecting BHK cells with RNA transcribed in vitro from cDNA clones and assaying the virus released into the culture

fluid after 48 h (Table 1). All nine mutations tested resulted in a higher yield of the chimera; between 2- and 20-fold more virus was produced than by the Ile-containing parental chimera. Although changes of twofold or so are marginal, it is of interest that in no case tested was the yield reduced. It is also of note that Val-380, one of the two changes found in faster-growing variants, gave only a threefold enhancement in yield and that yet this change was nevertheless selected during passage. Selection may be due in part to synergistic effects obtained with multiple changes (see Discussion). Changes of 5- to 20-fold are clearly significant and in line with the enhancement in yield induced by single mutations obtained in previous studies (7, 22). Ser-380, the second mutation seen in the passaged variants, produced about 11-fold more virus than did the parental chimera. Thr-380 and Asn-380 resulted in virus that

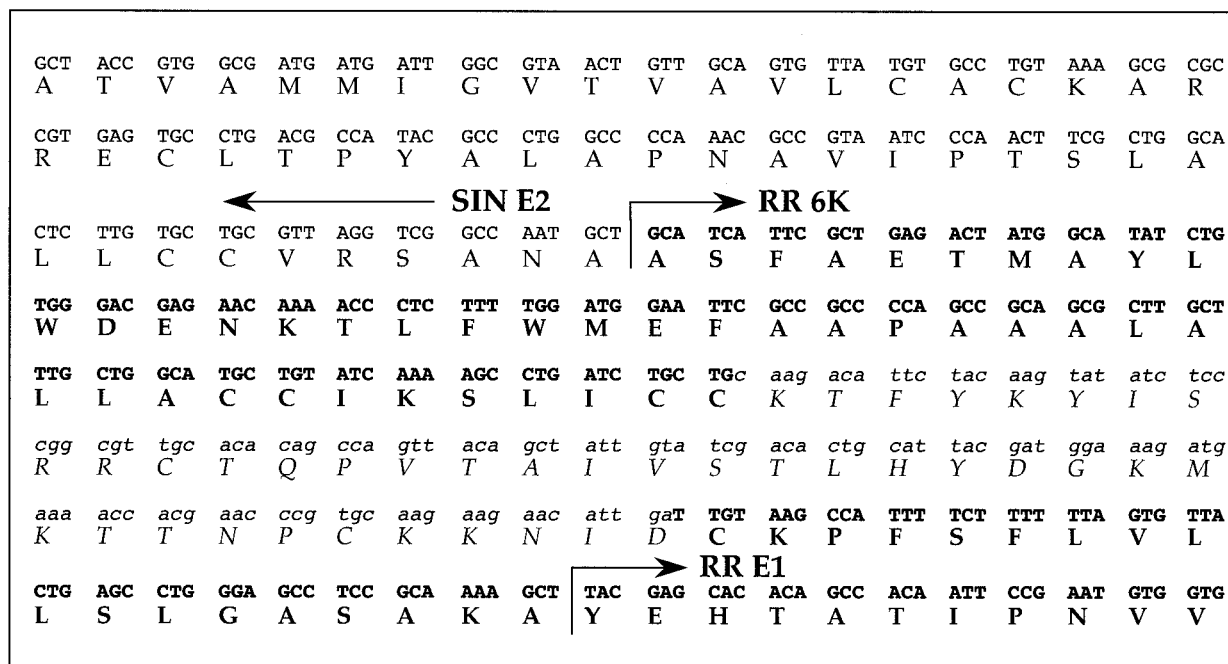


FIG. 2. Insertion in the 6K protein during passage of SIN(RRE1). The insertion into 6K that occurred in passage series 2 is shown. SIN sequences are shown in plain type; RR sequences are shown in boldface. The insert, shown in lowercase and italics, consists of nt 2594 to 2710 inclusive (117 nt) from SIN, corresponding to amino acids 306 to 344 of nsP2 and translated in the same frame as nsP2. The only change is that the last amino acid is Asp rather than Glu.

grew about the same as did virus with the Ser-380 mutation. The aromatic residues His and Tyr were even more efficient and resulted in 20-fold more virus than the parental chimera. The most surprising result was with Asp-380, which resulted in a 15-fold increase in virus production, even though putting a charge in the middle of a membrane anchor might have been expected to be deleterious.

A modest increase (≤ 20 -fold) in virus titer resulting from these mutations is consistent with the finding that more than one mutation is (usually) necessary to achieve the 100-fold increase in virus production observed upon selection of variants by continued passage and with previous results in which the changes in E2 or E1 have multiplicative effects upon virus yield (7, 22).

Growth curves of chimeras containing substitutions at Ile-380. In order to more carefully examine the effects of the mutations at Ile-380 upon the growth of the SIN(RRE1) chimera, BHK cells were transfected with RNA from selected mutants and samples were taken at intervals to determine the kinetics of release of virus (Fig. 3). All of the mutants tested, as well as the parental chimera, exhibited a pronounced delay in the production of virus. The Thr-380 mutant was delayed about 12 h in virus production, whereas the parental chimera was delayed by about 24 h. Other mutants showed intermediate delays. The extent of the delay and the ultimate titer produced were correlated, so that viruses with shorter delays produced more virus. The extended delay in virus production followed by a continuing rise in titer of the chimeric viruses at late times may be due to the appearance of new variants that are better adapted and that overgrow the culture (see references 7 and 22). Thus, the yields at early times are probably

more representative of the growth rates of the various chimeras than are the later yields.

The effect of mutations at Ile-380 upon SIN. The different mutations at Ile-380 were also moved into a SIN cDNA clone, and the effect of these mutations upon the growth of SIN was tested. Virus titers at 48 h following transfection with in vitro-transcribed RNA are shown in Table 1, and growth curves for selected mutants are shown in Fig. 3. Most of the mutations

TABLE 1. Growth of mutants at Ile-380^a

Residue at E2-380	Growth for:		
	SIN(RRE1)	SIN(RR6K)	SIN
Ile	(1)	(1)	(1)
Leu	2.1	0.8	1.0
Val	2.6		1.2
Ser	11		1.6
Thr	12	0.005	0.8
His	20	0.1	0.4
Tyr	20	0.3	0.1
Cys	5	1.0	1.3
Asn	12	0.04	0.4
Asp	15	0.0003	0.01

^a BHK cells were transfected with RNA transcribed from cDNA clones of the various mutants. Virus present in the culture fluid at 48 h after transfection was titered by plaque assay on BHK cells. The results are geometric averages from two to four experiments, except some of the SIN(RR6K) mutants, which were tested only once, and the Asp mutant in SIN(RRE1), which was tested 16 times.

The wild-type amino acid at position 380 of SIN E2 is Ile. Boldfaced amino acids and values indicate constructs made by swapping in restriction fragments from subclones of passaged virus. All other constructs were generated by site-specific mutagenesis. All yields are expressed as the ratio of the yield from a construct containing a mutant amino acid to the yield from the same construct containing Ile, which has been set arbitrarily to 1 (in parentheses).

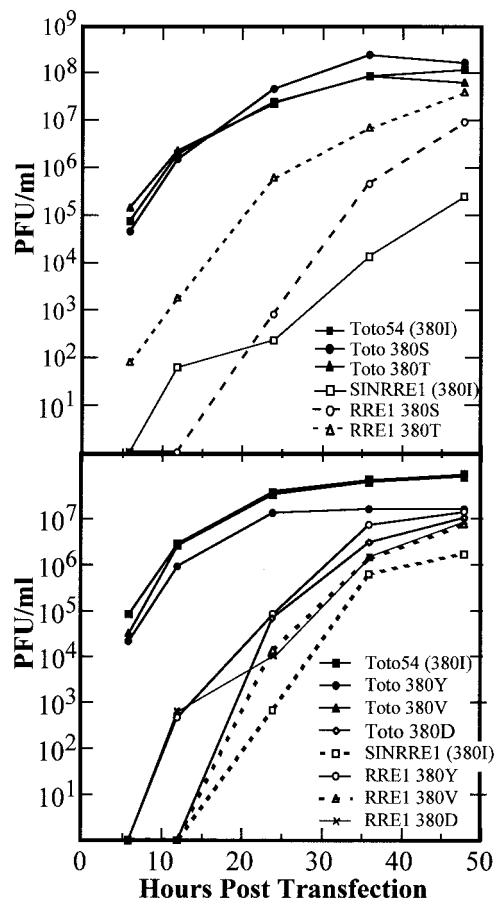


FIG. 3. Growth curve of Ile-380 mutants. Constructs labeled Toto 54 (or Toto) contained different amino acids at position E2 and amino acid 380 in the SIN background; constructs labeled SINRRE1 (or RRE1) had I380 mutations in the chimeric background. BHK cells were transfected with RNA transcribed *in vitro*, and samples were removed at intervals and titered for released virus by plaque assay on monolayers of BHK cells.

were neutral, with virus yields within a factor of about 2 of the wild-type (Ile-380) yield and with growth curves similar to that of the wild type. The two exceptions were Tyr-380, which led to a 10-fold decrease in virus yield, and Asp-380, which reduced the yield 100-fold. The effect of Asp-380 is not surprising, given that this change occurs within a hydrophobic anchor and that a charged residue at this position would be expected to affect the efficiency of the anchor and perhaps the rate of synthesis and transport of the glycoproteins. The surprise, in fact, is that this change enables the chimera to produce more virus (but not at wild-type levels). The finding that most changes in E2 that allow the chimeric virus to grow better have only modest effects upon the growth of SIN is consistent with earlier results in which we have found that changes in chimeric viruses often produce nonreciprocal effects.

Effects of Ile-380 mutations in SIN(RR6K). We have also studied a chimera called SIN(RR6K), which consists of SIN in which the 6K gene only has been replaced by that of RR. This chimera was found to be somewhat attenuated, producing about 10% as much virus as SIN, and studies of it suggested that 6K was involved in the maturation of the E2E1 hetero-

dimer (23). The various amino acid mutations at Ile-380 were also moved into the cDNA clone containing the SIN(RR6K) genome, and the growth of this chimera containing the various mutations was ascertained as done before (Table 1). The results were unexpected. Several of the mutations that led to increased growth of SIN(RRE1), in which both 6K and E1 are derived from RR, but which were neutral in SIN, were deleterious for SIN(RR6K), in which only the 6K is derived from RR. Thr-380 and Asn-380, and, to a lesser extent, His-380 fall into this category and led to a 10-fold-or-greater reduction in the yield of SIN(RR6K). Asp-380 is deleterious for SIN(RR6K), as it is for SIN.

DISCUSSION

Adaptation of SIN E2 to RR E1. Chimeric viruses are a powerful tool for the study of the assembly of alphaviruses. In a chimeric virus, the individual components, in this case glycoproteins RR E1, RR 6K, and SIN E2, are each known to be fully functional in their native contexts. Thus, when a chimera such as SIN(RRE1), containing RR E1, RR 6K, and SIN E2, is drastically attenuated, it is because of faulty protein-protein interactions, not because of misfolded or nonfunctional proteins. We have now identified 13 different changes in SIN E2 (involving 11 different residues), five different changes in RR E1, and an insertion in RR 6K that arose during passage of SIN(RRE1), and all of the changes that have been tested have been shown to adapt the disparate proteins to one another. Furthermore, individual changes have synergistic or multiplicative effects. In general, single changes result in about a 10-fold increase in virus production, two changes result in up to 100-fold more virus, and three changes result in up to 400-fold more virus. Interestingly, none of the changes examined had a significant effect upon the growth of the parental virus, whether RR or SIN.

The locations of the adaptive mutations identify regions of the glycoproteins that are important for their interactions. In the case of E1, four out of five mutations occurred in the stem-anchor regions, either within or proximal to the lipid bilayer. The fifth change occurred in domain III, which is positioned external to the stem-anchor I region. In the case of E2, the anchor region was also shown to be important for interaction with E1 by the occurrence of three variants with changes in Ile-380, which is found in the middle of the transmembrane domain. In addition, there was a pronounced cluster of mutations between residues 237 and 248 and a scattering of changes in the amino-terminal region from residues 72 to 150.

Although no atomic resolution structure of an alphavirus virion exists, cryoelectron microscopy has been used to determine the structure of several alphaviruses to resolutions of 9 to 25 Å (1, 13, 14, 24). The recent determination of the structure of E1 of Semliki Forest virus to near-atomic resolution (10), together with the determination of the positions within the virion of the two carbohydrate chains each on E1 and E2 of SIN (17), has allowed the positioning of E1 within the cryoelectron microscopy density and the crude positioning of E2. E1 is an elongated molecule that forms a layer apposed to the lipid bilayer called the skirt and that projects upward at an angle of about 50° to form the lower part of the spike. The two

carbohydrate chains of SIN E1 (at residues 139 and 245) are located only about 40 Å from the lipid bilayer (17). E2 projects the full length of the spike with residue 196, a carbohydrate attachment site, near the apex of the spike. Positioning of E3 (16, 20), the N-terminal domain of the precursor to E2 called PE2 or p62, at the most distal portion of the spike in Semliki Forest virus also suggests that the N terminus of E2 is near the apex of the spike. In contrast, residue 318 of E2, a carbohydrate attachment site, is located near the lipid bilayer. The transmembrane anchors of E1 and E2 are closely apposed to one another in the virion (13), winding around one another in a coiled-coil configuration (25). These structural studies support the molecular genetic evidence reported here, which reveals that interactions in the stem-anchor regions are an important component of the interactions between E1 and E2 that result in the formation of heterodimers.

Role of the 6K protein. The 6K protein is a small (55 residues in SIN, 60 residues in RR) hydrophobic peptide positioned between E2 and E1 in the sequence of the polyprotein precursor. This peptide is known to be important for virus assembly (3, 4, 6, 11) and to interact with E2 and E1 to promote the proper folding of the heterodimer (23). This interaction is specific because replacement of SIN 6K in SIN by RR 6K results in an altered conformation of the E2E1 heterodimer and decreased virus production (23). We assume that the 39-residue insertion into RR 6K in the chimera allows it to interact more efficiently with SIN E2. Furthermore, the finding that some changes in E2 Ile-380 had different effects in SIN and in SIN(RR6K), which differ only in the source of the 6K gene, suggests that the 6K protein interacts with the anchor domain of E2.

The role of the E2 anchor. Our results show that the sequences of the membrane anchors are important for the interaction of E1 and E2 and for their interaction with the 6K protein. The aligned sequences of the hydrophobic anchors of glycoprotein E2 from a number of different alphaviruses are shown in Fig. 4. These anchors show only limited sequence conservation, far less than do other domains of the glycoproteins. In particular, SIN E2 and RR E2 share 42% amino acid sequence identity over the entire E2 region but share only 19% identity in the anchor sequence. This relative lack of conservation presumably results because the key characteristic of the anchor is its hydrophobicity, which serves to anchor E2 within the membrane, and because the alpha-helical structure within the membrane is not required to form a precise structure. The results here, however, show that the sequence is not random and that the interaction of E1 and E2 requires that the anchor sequences must be capable of interaction.

It is of interest that position 380 shows little conservation. It is isoleucine in SIN but alanine, valine or leucine in other alphaviruses. It is leucine in RR and serine in Semliki Forest virus, and it is noteworthy that the change of Ile-380 to leucine had little effect upon the growth of the chimera, whereas the change to serine was selected during passage. It is also of interest that the bulky ring-containing amino acid histidine or tyrosine adapted the chimera even more efficiently than the selected changes but that these amino acids occur nowhere in any E2 anchor sequence except near the N-terminal end of the anchor of SIN (Tyr) and Whataroa (His) viruses. The nonspecific nature of the changes at Ile-380 that result in better virus

						L, T, H, Y, C, N, D, S, V				
		365		375				385		
SIN	HRHPV	Y	TILAVASAT	V	A	I	GVTVA	VLCACKAR		
AURA	HLYPF	Y	TVTIVLSGMG	L	A	I	GLVIS	ILCCCKAR		
RR	GLYPA	A	TIAAVSGAS	L	MAL	L	TLAAT	CCMLATAR		
SF	GLYPA	A	TVSAVVGMS	L	LAL	I	SIFAS	CYMLVAAR		
MID	GLYPA	T	VAVCVGLA	C	VIL	L	ALSAS	CCLCVSAR		
BF	DLYPY	W	TITVLASLG	L	LIV	I	SSGFS	CFLCSVAR		
MAY	GLHPT	T	TIVVVIRVS	V	VVL	L	SFAAS	VYMCVVAR		
SAG	GLYPA	A	TIAAVSAAG	L	AVV	L	SLLAS	CYMFATAR		
ONN	ELYPT	T	TIAVLAAS	I	VIT	S	LVGLS	LGMICAR		
WEE	HRHPV	Y	TIVIVLCGVA	L	AII	V	GTASS	AACIAKAR		
EEE	NRYPL	T	TIIIGLCTCV	A	IIM	V	SCVTS	VWLLCRTR		
VEE	HRYPM	S	TILGLSICA	A	IIV	V	SVAAS	TWLFCKSR		
SES	DKHPL	L	TASVITVTA	L	LTV	I	GFALA	AWLICAR		
SPDV	KHHPE	Y	AWAFVGVAC	G	L	I	AACMF	ACACSRVR		
SLDV	KHHPE	Y	AWAFVGVAC	G	L	V	AACMF	ACACNRVR		

FIG. 4. Transmembrane anchor of E2. Sequences of the E2 anchor sequences for several alphaviruses are compared. AURA is Aura virus, SF is Semliki Forest virus, MID is Middelburg virus, BF is Barmah Forest virus, MAY is Mayaro virus, SAG is Sagiyama virus, ONN is O'Nyong-nyong virus, VEE is Venezuelan equine encephalitis virus, SES is Southern elephant seal louse virus, SPDV is salmon pancreas disease virus, and SLDV is rainbow trout sleeping disease virus.

yield in the chimera suggests that the failure of RR E1 and SIN E2 to interact properly might arise from steric hindrance from the bulky Ile residue. The fact that most of these changes are neutral in SIN could result from an ability to accommodate a smaller residue into the SIN structure.

Studies of Western equine encephalitis virus (WEE) suggest that the anchors might interact with the nucleocapsid as well. WEE is a naturally occurring chimera that arose by recombination between Eastern equine encephalitis virus (EEE) and SIN, or an ancestor of SIN, more than 1,000 years ago (5, 21). In WEE, the capsid protein and the N-terminal half of E3 are derived from EEE, whereas the C-terminal half of E3, as well as of E2 and E1, are derived from the SIN ancestor. In the chimera a number of amino acid substitutions have occurred that appear to adapt these disparate elements to one another. In particular, as in the case studied here, six amino acid substitutions occurred within the WEE E2 anchor in which the amino acid switched from that in SIN to that found in EEE (Fig. 4). Similarly a number of amino acid changes occurred in the E1 anchor. The apparent selection of these changes suggests that the sequence of the anchor is important for the interaction of the glycoproteins with the nucleocapsid. Thus, the anchors appear to serve specific functions in virus assembly that are not limited to simple anchorage of the glycoproteins in the envelope of the virion.

ACKNOWLEDGMENTS

We acknowledge Jian-Sheng Yao for helpful discussions and participation during the initial phases of this project.

This work was supported by grants AI 10793 and AI 20612 from the National Institutes of Health.

REFERENCES

1. Cheng, R. H., R. J. Kuhn, N. H. Olson, M. G. Rossmann, H.-K. Choi, T. J. Smith, and T. S. Baker. 1995. Nucleocapsid and glycoprotein organization in an enveloped virus. *Cell* **80**:621–630.
2. Forsell, K., L. Xing, T. Kozlovska, R. H. Cheng, and H. Garoff. 2000. Membrane proteins organize a symmetrical virus. *EMBO J.* **19**:5081–5091.
3. Gaedigk-Nitschko, K., M. Ding, and M. J. Schlesinger. 1990. Site-directed mutations in the Sindbis virus 6K protein reveal sites for fatty acylation and the underacylated protein affects virus release and virion structure. *Virology* **175**:282–291.
4. Gaedigk-Nitschko, K., and M. J. Schlesinger. 1990. The Sindbis virus 6K protein can be detected in virions and is acylated with fatty acid. *Virology* **175**:274–281.
5. Hahn, C. S., S. Lustig, E. G. Strauss, and J. H. Strauss. 1988. Western equine encephalitis virus is a recombinant virus. *Proc. Natl. Acad. Sci. USA* **85**:5997–6001.
6. Ivanova, L., S. Lustig, and M. J. Schlesinger. 1995. A pseudo-revertant of a Sindbis virus 6K protein mutant, which corrects for aberrant particle formation, contains two new mutations that map to the ectodomain of the E2 glycoprotein. *Virology* **206**:1027–1034.
7. Kim, K. H., E. G. Strauss, and J. H. Strauss. 2000. Adaptive mutations in Sindbis virus E2 and Ross River virus E1 that allow efficient budding of chimeric viruses. *J. Virol.* **74**:2663–2670.
8. Kuhn, R. J., D. E. Griffin, K. E. Owen, H. G. M. Niesters, and J. H. Strauss. 1996. Chimeric Sindbis-Ross River viruses to study interactions between alphavirus nonstructural and structural regions. *J. Virol.* **70**:7900–7909.
9. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**:488–492.
10. Lescar, J., A. Roussel, M. W. Wein, J. Navaza, S. D. Fuller, G. Wengler, G. Wengler, and F. A. Rey. 2001. The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primer for fusogenic activation at endosomal pH. *Cell* **105**:137–148.
11. Liljeström, P., S. Lusa, D. Huylebroeck, and H. Garoff. 1991. In vitro mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. *J. Virol.* **65**:4107–4113.
12. Lopez, S., J.-S. Yao, R. J. Kuhn, E. G. Strauss, and J. H. Strauss. 1994. Nucleocapsid-glycoprotein interactions required for alphavirus assembly. *J. Virol.* **68**:1316–1323.
13. Mancini, E. J., M. Clarke, B. E. Gowen, T. Rutten, and S. D. Fuller. 2000. Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol. Cell* **5**:255–266.
14. Paredes, A., K. Alwell-Warda, S. C. Weaver, W. I. Chiu, and S. J. Watowich. 2001. Venezuelan equine encephalomyelitis virus structure and its divergence from Old World alphaviruses. *J. Virol.* **75**:9532–9537.
15. Paredes, A. M., D. T. Brown, R. B. Rothnagel, W. Chiu, R. J. Schoepf, R. E. Johnston, and B. V. V. Prasad. 1993. Three-dimensional structure of a membrane-containing virus. *Proc. Natl. Acad. Sci. USA* **90**:9095–9099.
16. Paredes, A. M., H. Heidner, P. Thuman-Commike, B. V. V. Prasad, R. E. Johnston, and W. Chiu. 1998. Structural localization of the E3 glycoprotein in attenuated Sindbis virus mutants. *J. Virol.* **72**:1534–1541.
17. Pletnev, S. V., W. Zhang, S. Mukhopadhyay, B. R. Fisher, R. Hernandez, D. T. Brown, T. S. Baker, M. G. Rossmann, and R. J. Kuhn. 2001. Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. *Cell* **105**:127–136.
18. Strauss, J. H., and E. G. Strauss. 1994. The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**:491–562.
19. Strauss, J. H., E. G. Strauss, and R. J. Kuhn. 1995. Budding of alphaviruses. *Trends Microbiol.* **3**:346–350.
20. Vénien-Bryan, C., and S. D. Fuller. 1994. The organization of the spike complex of Semliki Forest virus. *J. Mol. Biol.* **236**:572–583.
21. Weaver, S. C., W. Kang, Y. Shirako, T. Rumenapf, E. G. Strauss, and J. H. Strauss. 1997. Recombinational history and molecular evolution of Western equine encephalomyelitis complex alphaviruses. *J. Virol.* **71**:613–623.
22. Yao, J., E. G. Strauss, and J. H. Strauss. 1998. Molecular genetic study of the interaction of Sindbis virus E2 with Ross River E1 for virus budding. *J. Virol.* **72**:1418–1423.
23. Yao, J. S., E. G. Strauss, and J. H. Strauss. 1996. Interactions between PE2, E1, and 6K required for assembly of alphaviruses studied with chimeric viruses. *J. Virol.* **70**:7910–7920.
24. Zhang, W., B. R. Fisher, N. H. Olson, J. H. Strauss, R. J. Kuhn, and T. S. Baker. 2002. Aura virus structure suggests that the T=4 organization is a fundamental property of viral structural proteins. *J. Virol.* **76**:7239–7246.
25. Zhang, W., S. Mukhopadhyay, S. V. Pletnev, T. S. Baker, R. J. Kuhn, and M. G. Rossmann. Placement of the structural proteins in Sindbis virus. *J. Virol.*, in press.
26. Zoller, M. J., and M. Smith. 1984. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. *DNA* **3**:479–488.