Frequent Frameshift and Point Mutations in the SH Gene of Human Metapneumovirus Passaged In Vitro

Stéphane Biacchesi, Brian R. Murphy, Peter L. Collins and Ursula J. Buchholz

Published Ahead of Print 21 March 2007.

Updated information and services can be found at:
http://jvi.asm.org/content/81/11/6057

REFERENCES

These include:
This article cites 29 articles, 15 of which can be accessed free at: http://jvi.asm.org/content/81/11/6057#ref-list-1

CONTENT ALERTS

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»
Human metapneumovirus (HMPV), which is first described in 2001 (26), is a member of the Paramyxoviridae family and has been assigned to the Metapneumovirus genus of the Pneumovirinae subfamily, together with avian metapneumovirus. Human respiratory syncytial virus (HRSV), the member of the Pneumovirinae that has been studied in the greatest detail, belongs to the second genus of Pneumovirinae, Pneumovirus (10). HMPV is an important cause of respiratory tract infection worldwide, resembling HRSV, and is estimated to be associated with 5 to 15% of pediatric hospitalizations due to respiratory tract infections (13, 18, 27, 28). There are currently no licensed vaccines or specific therapies for HMPV (9).

Like all members of the Paramyxoviridae family, HMPV is an enveloped, single-stranded, negative-sense RNA virus. The genome is approximately 13 kb in length, contains eight genes in the order 3'-N-P-M-F-M2-SH-G-L-5', and encodes nine proteins, with the M2 mRNA containing two overlapping open reading frames (ORFs) that are expressed as two separate proteins, M2-1 and M2-2 (4, 7, 25). The HMPV proteins are classified into nine proteins, whereas G and SH appear to have minor and insignificant roles, respectively, in inducing neutralizing antibodies and protection (23). We and others previously developed reverse-genetics systems for HMPV (5, 15), whereby complete infectious virus can be recovered from cultured cells that have been transfected with plasmids encoding a positive-sense copy of the viral genome and the N, P, M2-1, and L proteins. This recombinant system provides the basis for introducing defined mutations into infectious HMPV. In particular, this approach is currently being used to develop attenuated mutants of HMPV for use in a live intranasal vaccine (2, 9, 21). Pertinent to the present report, studies with recombinant viruses showed that the deletion of SH and G, individually or in combination, had little or no apparent effect on replication in cell culture (6). Virus in which the M2-1 and M2-2 ORFs were silenced separately or in which the entire gene was deleted also replicated in cell culture with an efficiency similar to that of wild-type HMPV, although the deletion of M2-1 appeared to be slightly attenuating (7).

The present report describes how, in the course of preparing recombinant derivatives of the CAN97-83 clinical isolate of human metapneumovirus (HMPV), consensus nucleotide sequencing of the recovered RNA genomes provided evidence of frequent sequence heterogeneity at a number of genome positions. This heterogeneity was suggestive of sizable subpopulations containing mutations. An analysis of molecularly cloned cDNAs confirmed the presence of mixed populations. The biologically derived virus on which the recombinant system is based also contained sizeable mutant subpopulations, whose presence was confirmed by biological cloning and nucleotide sequencing. Most of the mutations occurred in the SH gene. For example, partial consensus sequencing of 40 independent preparations of recombinant HMPV (wild-type and various derivatives) showed that 31 of these preparations contained a total of 41 instances of small insertions in the SH gene and a total of five small insertions elsewhere. In each of these 31 preparations, there was at least one insert in SH that changed the reading frame and would yield a truncated protein. Nearly all of these insertions involved adding one or more A residues to various tracks of four or more A residues, with the most frequent site being a tract of seven A residues. There were also two instances of nucleotide deletions and numerous instances of nucleotide substitution point mutations, mostly in the SH gene. The occurrence of mutant subpopulations was greatly reduced by the replacement of the SH gene with a synthetic version in which these oligonucleotide tracts were eliminated by silent nucleotide changes. We suggest that we frequently detected subpopulations in which the expression of full-length SH protein was ablated because it provided a modest selective advantage to this clinical isolate in vitro. Adaptation involving the functional loss of a gene is unusual for an RNA virus.
in which the expression of SH has been affected suggests that this might confer a subtle selective advantage in vitro.

MATERIALS AND METHODS

Cells. LLC-MK2 rhesus monkey kidney cells (ATCC CCL 7.1) or Vero African green monkey (AGM) kidney cells (ATCC CCL 81) were maintained in OptiMEM I (Invitrogen) supplemented with 5% fetal bovine serum and 2 mM l-glutamine or in OptiPro SFM (Invitrogen) supplemented with 4 mM l-glutamine, respectively. BSR T7/5 cells are baby hamster kidney 21 cells that constitutively express T7 RNA polymerase (8). They were maintained in Glasgow modified essential medium supplemented with 2 mM l-glutamine and amino acids (1× final modified essential medium amino acids solution; Invitrogen), 10% fetal bovine serum, and 1 mg/ml of Geneticin.

Virus propagation and quantitation. The Canadian clinical isolate CAN97-83 (19) and the various recombinant HMPVs (rHMPVs) were propagated in LLC-MK2 or Vero cells at 32°C in the absence of serum and in the presence of trypsin as previously described (5). Virus titers were determined by plaque assay in Vero cells under a methylcellulose overlay containing trypsin as described previously (5).

Recombinant viruses. We previously constructed plasmid pHMPV, which encodes a copy of the antigenic RNA of HMPV CAN97-83 (5) that is identical to the CAN97-83 consensus sequence (GenBank accession number AY297749) except for four nucleotide substitutions in the M-F intergenic region that create a unique NheI site. This plasmid was modified by the addition of three nucleotide substitutions in the M2-SH intergenic region that create a BsiWI site and three nucleotide substitutions in the SH-G intergenic region that create a BsrGI site, resulting in pHMPV(BsiWI/BsrGI). All of the recombinant viruses in this study were in this backbone.

A “stabilized” version of the SH gene (SHs) was created by modifying the sequence of the SH gene to replace 31 A and T residues contained within homopolymer tracts with C or G without changing the amino acid coding (Fig. 1). The SHs cDNA was synthesized commercially (Blue Heron). To exchange the naturally occurring SH gene with SHs, the NheI/Acc65I fragment of pHMPV(BsiWI/BsrGI) containing the F, M2, SH, and G genes was subcloned and the SH gene was removed by BsiWI/BsrGI digestion. The synthetic SHs gene was cloned into this window such that the only changes were the A/T to C/G substitutions shown in Fig. 1. The NheI/Acc65I fragment bearing SHs was replaced into pHMPV(BsiWI/BsrGI), creating the full-length antigenic plasmid pHMPV-SHs.

Prior to the recovery of recombinant virus, the sequence of each fragment that had been subjected to mutagenesis was completely confirmed by nucleotide sequencing performed with an ABI 3730 automatic sequencer using the BigDye terminator ready reaction kit version 1.1 (Applied Biosystems) and specific primers. Recombinant viruses were recovered as described previously (3, 5). Sequence analysis of viral RNA was carried out as described previously (3). In some cases, as indicated in the text, these reverse transcription-PCR (RT-PCR) products were gel purified and cloned using a commercial blunt vector (Perfectly Blunt cloning kit; Novagen) and individual clones were sequenced.

Virus purification and Western blot assay. The indicated viruses were grown in Vero cells and harvested and purified as described previously (6), except that the clarified medium supernatants were subjected directly to centrifugation in sucrose gradients rather than first being pelleted overnight. The protein content of each preparation was quantified using a bicinchoninic acid protein assay kit (Pierce). As described previously (6), the samples were denatured, reduced, and subjected to electrophoresis on 4 to 12% Polyacrylamide Bis-Tris gels (NuPAGE; Novex Bis-Tris; Invitrogen) and the separated proteins were stained with Coomassie blue staining kit (Invitrogen) or subjected to Western blot analysis with a rabbit antiserum raised against peptide SH82-96, representing amino acids 82 to 96 of the SH protein.

Replication of rHMPV and rHMPV-SHs in hamsters. Twelve-week-old Golden Syrian hamsters in groups of 12 were inoculated intranasally under light methoxyflurane anesthesia with 0.1 ml of L15 medium containing 106 PFU of rHMPV or rHMPV-SHs. On days 3 and 5 postinfection, six animals from each group were sacrificed and their lungs and nasal turbinates were harvested. Tissue homogenates were prepared and analyzed for infectious virus by plaque assay.

Comparison of rHMPV and rHMPV-SHs in AGMs. Using previously described methods (2, 24), young adult AGMs were confirmed to be negative for serum-neutralizing antibodies to HMPV and were inoculated simultaneously by the intranasal and intratracheal routes with a 1-ml inoculum per site containing 106 PFU of rHMPV and rHMPV-SHs in L15 medium. A mock infection control group received L15 medium alone. Each group included six animals except for the mock infection group, which contained two animals. Clinical observations were made daily on days 0 to 12 following inoculation. Nasopharyngeal swabs were collected daily for 10 days following inoculation and on day 12. Tracheal lavage samples were collected on days 2, 4, 6, 10, and 12 postinfection. The amount of virus present in the samples was quantified by plaque assay. On day 28, each monkey was challenged by the intranasal and intratracheal routes with 107 PFU of rHMPV in a 1.0-ml volume per site. Virus shedding was examined by plaque assay of nasopharyngeal and tracheal lavage samples collected on days 2, 4, 6, and 8 postchallenge. Serum samples were collected on days 0, 28, and 56, and the titers of HMPV-neutralizing antibodies were determined by using an endpoint dilution neutralization assay as described previously (3).

RESULTS

Frequent mutations detected in recombinant HMPV. In the course of recovering various recombinant HMPVs using reverse genetics, consensus genome sequence analysis of preparations of recovered viruses revealed a surprisingly frequent occurrence of what appeared to be viral subpopulations containing adventitious mutations, especially in the SH gene (Fig. 1 and Table 1). Figure 2A shows an example involving wild-type HMPV that was recovered by transfection (passage level 0 [P0]) and passed serially in LLC-MK2 cells, with an initial blind passage (P1) and subsequent passages at an input multiplicity of infection (MOI) of 0.01 PFU/cell. Total intracellular RNA was isolated from passages P2, P3, and P4, and RT-PCR was used to amplify a genome fragment of 4,781 bp containing the end of the M gene, the entire F, M2, SH, and G genes, and the beginning of the L gene (nucleotides positions 2719 to 7499 in the HMPV genome). This fragment was then sequenced in its entirety directly from the RT-PCR product. As shown in Fig. 2A, sequence heterogeneity was apparent in P4, involving what appeared to be the presence of a mixed-base assignment at SH gene position 78, which in the unmodified sequence follows a run of seven A residues (all sequence descriptions are positive sense unless otherwise noted). Thus, position 78, which normally is a T residue immediately downstream of the A tract, appeared to be a mixture in which the predominant peak was T and the minor peak was A.

In order to estimate the abundance of the subpopulation at position 78, we constructed two cDNA clones of this region, one containing the wild-type sequence and another with an A insertion added to the 7A tract. The cDNAs were mixed in a graded series of ratios and used as templates for PCR, followed by gel purification of the specific band and nucleotide sequencing. The resulting electropherograms based on these standards of known ratios (data not shown) were used to estimate ratios in mixed peaks from the infected-cell RNA. In the case of the rHMPV electropherogram shown in Fig. 2A, the minor population in P4 was thus estimated to be approximately 20% of the total. We also performed RT-PCR and nucleotide sequencing of RNA purified from the clarified medium supernatant of P4, thus representing virion-associated RNA. This exhibited the same pattern of apparent heterogeneity at this position, with the minor A peak representing approximately 20% of the total (data not shown). Thus, intracellular and virion-associated RNA contained the same apparent mixture of species.

As a second example, Fig. 2B illustrates results with a virus preparation of a mutant from which the entire G gene was deleted (rHMPV-ΔG). In this case, sequence analysis of intracellular RNA from P2 to P5 revealed the same apparent heterogeneity at position 78. In this particular experiment, the A
peak increased progressively from 5% in P2 to 70% in P5, thus becoming the predominant peak. In the electropherograms in which this secondary peak became particularly abundant, it was evident that there also were secondary peaks at the positions to the right, consistent with this result representing a population of molecules in which the run of seven A residues had been increased by the insertion of an additional A residue.

The SH protein is a transmembrane virion envelope protein of 179 amino acids (for this strain) that contains a combined signal sequence and membrane anchor near the amino terminus.
TABLE 1. Summary of the mutations observed in 40 independently derived preparations of various recombinant HMPVs

<table>
<thead>
<tr>
<th>No. of preparations</th>
<th>Insertion(s) in the SH gene</th>
<th>Deletion in the SH gene</th>
<th>Point mutation(s) in the SH gene</th>
<th>Insertion(s) elsewhere</th>
<th>G ORF point mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2*</td>
<td>A39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3*</td>
<td>A57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/</td>
<td>A78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78</td>
<td></td>
<td>4/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78</td>
<td></td>
<td>11/9</td>
<td>3/1</td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 (X2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 (X4)</td>
<td></td>
<td>9/6 and 1/ATG</td>
<td>1/1</td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 (X5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A300 (X2)</td>
<td></td>
<td>10/8</td>
<td>M2 GE</td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A407</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 + A78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 + A184</td>
<td>A52*</td>
<td>1/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 + T397</td>
<td></td>
<td>3/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 + A575</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 (X3) + A300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 (X3) + A300</td>
<td></td>
<td>33/18 and stop + 4*</td>
<td>L ORF + 16/</td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>T397 + A460</td>
<td></td>
<td>2/2</td>
<td>F GE*; L NC*</td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 + A258</td>
<td>A35*</td>
<td>3/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>A78 + A210 + T624</td>
<td></td>
<td>1/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9*</td>
<td>No mutations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The virus preparations are grouped according to the types of mutations that they contained and are listed in order of increasing position in the SH gene and number of insertions plus deletions. The specific identities of the viruses are not indicated for the sake of simplicity. The mutations were identified in a consensus sequence of a 4,781-bp RT-PCR fragment spanning from the end of M (the primer represented nucleotide positions 2719 to 2736) to the beginning of L (nucleotide positions 7409 to 7479), including the complete F, M2, SH, and G genes, unless otherwise noted (see footnotes f and k). Nucleotide assignments are positive sense.

| f | Insertions were named according to the inserted nucleotide and the nucleotide sequence position. X2, X3, X4, and X5 refer to the insertion of 2, 3, 4, and 5 A residues, respectively, at the indicated positions.

| g | Deletions were named according to the deleted nucleotide and the nucleotide sequence position.

| h | Point mutations were annotated as the number of nucleotide substitutions/number of amino acid substitutions, except as noted.

| i | * at least one preparation was also analyzed by molecular cloning and sequencing.

| j | The consensus sequence was determined for the complete genome; a complete sequence was determined for one of the six preparations with SH insertion A78 and for two out of nine preparations with no mutations in SH.

| k | The insertion was of a single A residue at antigenome position 4547 into the oligo(A) tract that comprises the end of the M GE signal.

| l | Point mutations in the G ORF were annotated as the number of nucleotide substitutions/number of amino acid substitutions.

| m | * 1/ATG is a point mutation that ablated the SH ORF start codon (ATG to AAG).

| n | Molecular cloning showed that the A78 insertion and A52 deletion in this virus preparation represent two distinct viral subpopulations; each creates a frameshift in the SH ORF.

| o | The consensus sequence was from the end of M2 (the primer represented nucleotide positions 4955 to 4976) to the beginning of L (nucleotide positions 7499 to 7479).

| p | The insertion was of two A residues at antigenome position 4698 into the oligo(A) tract that comprises the end of the F GE signal.

| q | The consensus sequence was determined for the complete genome; a complete sequence was determined for one of the six preparations with SH insertion A78 and for the remaining five preparations, a 2,545-bp fragment that includes the end of the F gene, the SH and G genes, and the start of the L gene.

| r | For the sake of simplicity, these are not identified individually, since the incidence of adventitious mutations did not seem to be affected by the identity of the virus (data not shown).

| s | Nine of these preparations were based on viruses that lacked the G gene; otherwise, all of the genes were present.

| t | Consensus RNA sequences were determined for the following genome regions: for 31 preparations, the above-mentioned 4,781-bp fragment (36% of the genome) that spans from the end of the M gene to the beginning of the L gene; for four preparations, a 2,545-bp fragment that includes the end of the M2 gene, the SH and G genes, and the start of the L gene (positions 4955 to 7499 in the HMPV genome); and for the remaining five preparations, the complete genome (Table 1).

| u | The preparations had been passaged two to four times in LLC-MK2 or Vero cells following recovery; the choice of cell

| v | Analysis of 40 independent preparations of recombinant virus. We went on to examine a total of 40 recombinant virus preparations from independent reverse genetic recoveries, including the two examples already mentioned (Table 1). The viruses that were examined included several independent preparations of the wild type as well as a number of other derivatives. For the sake of simplicity, these are not identified individually, since the incidence of adventitious mutations did not seem to be affected by the identity of the virus (data not shown). Nine of these preparations were based on viruses that lacked the G gene; otherwise, all of the genes were present. Consensus RNA sequences were determined for the following genome regions: for 31 preparations, the above-mentioned 4,781-bp fragment (36% of the genome) that spans from the end of the M gene to the beginning of the L gene; for four preparations, a 2,545-bp fragment that includes the end of the M2 gene, the SH and G genes, and the start of the L gene (positions 4955 to 7499 in the HMPV genome); and for the remaining five preparations, the complete genome (Table 1). The preparations had been passaged two to four times in LLC-MK2 or Vero cells following recovery; the choice of cell
line did not seem to make a difference and is not indicated for the sake of simplicity.

Of the 40 virus preparations, 31 had evidence of a substantial mutant subpopulation(s) as defined by the presence of one or more secondary peaks in the sequencing electropherograms at an estimated abundance of 5 to 10% or greater (Table 1). The most common mutation was the insertion of one nucleotide (and up to five nucleotides) into homo-oligomer tracts already containing two to seven residues of the same identity. Each of the 31-mutant-containing preparations had at least one insertion, and a number had more than one, for a total of 46 instances. Of the total of 46 instances of insertion, 41 occurred in the SH gene, involving 12 different sites. Remarkably, each of the 31 preparations had at least one insert in the SH gene (eight preparations contained two inserts, and one preparation contained three) (Table 1 and Fig. 1). The result was that each of the 31 preparations sustained a frameshift in the SH ORF that would result in a truncated SH protein (Table 1 and Fig. 1). Two insertions (A575 and T624) (Fig. 1) occurred outside of the SH ORF, but each of these occurred together with at least one other insertion within the ORF. Where there were two or three inserts, in no case was the SH reading frame restored. In cases where an insert involved three nucleotides and thus added a codon but did not change the reading register, frameshifts disrupting the ORF occurred elsewhere (Table 1).

The other five instances of insertion occurred outside the SH gene and involved four different sites (Table 1). Three instances involved the insertion of A into the oligo(A) tract at the downstream end of the M2 or F gene end (GE) signal. This mutation probably would not have a major effect on the virus since the length of the oligo(A) tract differs from gene to gene and the altered GE signals did not exceed the lengths found in nature (4, 25). Another insertion occurred in the downstream noncoding region of the L gene and likely would be without effect. The remaining insertion occurred immediately following the last codon of the L ORF and created a frameshift that would add 16 codons from the alternate reading frame. Forty-two of the 46 instances of insertion occurred in oligo(A) or oligo(T) tracts of four to seven residues involving 12 different sites (SH positions 39, 57, 78, 184, 210, 300, 397, 575, and 624 and the M2 GE, F GE, and L ORF) (Table 1 and Fig. 1). Of the remaining four instances, two occurred in runs of three A residues (SH positions 407 and 460) and two occurred in runs of two A or T residues (SH position 258 and the L noncoding region). The most frequently used site (21 of 46 insertions) was at SH gene position 78, which was a run of seven

FIG. 2. Frameshift mutation occurring at position 78 of the SH gene after several passages in LLC-MK2 cells of two recombinant HMPV viruses: wild-type rHMPV (A) and rHMPV-ΔG (B). The electropherograms show the nucleotide sequence (SH gene positions 65 to 90) surrounding the site of insertion (black arrows). The passage number and the MOI (PFU/cell) in each passage are indicated on the left. Indicated on the right is the titer (PFU/ml) of each harvest, the day of harvest, and the approximate percentage of the mutant assignment in the preparation.
A residues in the unmodified gene and has been described above for the rHMPV and rHMPV-ΔG examples (Fig. 2).

Two of the 31 virus preparations also contained deletions of a single A residue that shifted the SH reading frame (Table 1 and Fig. 1). These deletions occurred at two different sites, A52 and A335, located in runs of five and two A residues, respectively (Table 1 and Fig. 1). These deletions at A52 and A335 occurred in preparations that also contained one or two inserts, respectively; in neither case did a combination occur that restored the SH ORF.

Ten of the 31 virus preparations in Table 1 also contained evidence of point mutations, of which 79 instances occurred in SH and 10 in G (Table 1). Some preparations contained one or a few scattered point mutations. Other preparations contained multiple point mutations that preferentially involved A-to-G or T-to-C transitions and might represent biased hypermutations generated by host cell adenosine deaminases (reviewed in reference 22). The 79 instances of point mutations in SH were nearly evenly divided between the first (n = 17), second (n = 22), and third (n = 19) codon positions. All 19 of the substitutions in the third position were silent at the amino acid level. In contrast, of the 39 substitutions in the first and second positions, nearly all resulted in changes at the amino acid level: 36 were missense, including 1 that removed the stop codon and extended the length of the SH protein by four amino acids and 1 that ablated the start codon (ATG to AGG), and only 3 were silent. Thus, changes in amino acid coding in SH were readily accepted. The 10 point mutations in G involved five substitutions in the third codon position, each of which was silent, and five in the first and second positions, each of which caused a nonsense mutation. Thus, G accumulated many fewer point mutations, with perhaps a somewhat lower frequency of changes at the amino acid level.

Confirmation in cloned cDNAs. We subjected RT-PCR products from 17 of the 31 mutation-containing preparations (identified in Table 1) to molecular cloning, and in each case, we sequenced five to eight cDNA clones (data not shown). The results (data not shown) confirmed that the insertions and deletions in any given preparation that were predicted based on peaks in the sequencing electropherograms indeed were present in a proportion of the cDNA clones derived from that preparation, with another proportion having the wild-type sequence. In instances where the examined preparation contained more than one insertion or deletion, each combination was present together in some of the cloned molecules. The only exception was the one preparation containing the A78 insert and A52 deletion, which were always segregated. Individually, the A78 insertion and A52 deletion each disrupted the SH ORF; the combination of the two together would have restored the SH reading frame with the introduction of five amino acid substitutions between the two frameshifts, but this did not occur. In some cases where more than one insertion or deletion was found together in individual cDNA clones, one of them also could be found individually in some cDNA clones.

The nucleotide point substitutions in the examined preparations also were confirmed in the cDNA clones. When multiple mutations were present in a population, they often occurred in various combinations in various individual cDNA clones. This indicated that they were present in multiple mutant subpopulations, thus representing a swarm of mutant molecules. In the cDNA clones from these various preparations, the SH gene insertions sometimes could be observed in the absence of point mutations, but the converse usually was not observed. These observations suggest that the gene insertion usually was acquired first, followed by the point mutations.

Frameshift and point mutations also occurred in biologically derived HMPV. We also examined the clinical HMPV CAN97-83 isolate, on which the recombinant system is based, for the presence of sequence heterogeneity. Sequence analysis of our laboratory stock, which in our hands had been passaged three times in LLC-MK2 cells, revealed a minor secondary peak of \(-5\%\) or less abundance at SH gene position 78. Two further passages in LLC-MK2 cells at any one of several input MOIs resulted in the emergence of an A78 peak that achieved an abundance of 10 to 40% (Fig. 3A). We performed molecular cloning of the P5 MOI 1 preparation (Fig. 3A, bottom) and sequenced individual cDNA clones, confirming the presence of the A78 insert in a proportion of the molecules as well as the presence of multiple point mutations.

We then plaque purified 12 clones from the P3 virus stock, amplified them once in LLC-MK2 cells, and subjected them to RT-PCR and nucleotide sequencing. As shown in Fig. 3B, each of these 12 individual plaques was heterogeneous at position 78, with 5 to 90% of the population appearing to contain the A78 insert. The inability to select clean populations may have been due to cross-contamination by diffusion between plaques during the long period of incubation necessitated by the slow growth of HMPV, although to some extent, it might also be due to ongoing insertion events.

We subjected the plaque-selected population with the greatest amount of A78 insertion (90%) to a second round of plaque purification, picked 22 plaques, amplified them once, and subjected them to RT-PCR and sequence analysis (Fig. 3B). Of the 22 preparations, 12 appeared to be homogeneous: 5 contained the wild-type SH sequence, 8 contained the A78 insertion, and 1 contained an insertion of two A residues (A78[X2]). The remaining 10 clones appeared to be heterogeneous at position 78. This heterogeneity might reflect insufficiency of plaque isolation, but it also might represent ongoing insertion. In any event, this result confirmed that the insertion of one or more A residues at position 78, the major mutation in the recombinant viruses, also occurs in biologically derived virus and indeed can be isolated in a biologically cloned preparation.

Construction of an evaluation of rHMPV bearing SHs. The amount of sequence variability observed in these experiments would be problematic for developing a live-attenuated HMPV vaccine and obtaining regulatory approval. Since most of the mutations occurred in the SH gene and frequently targeted tracts of four or more A or T residues, we designed an SH gene in which all of the A or T tracts of four residues or more as well as a number of three-A and three-T tracts occurring within the ORF were interrupted by introducing silent substitutions of C or G (Fig. 1). The naturally occurring SH gene was replaced with its stabilized version, and the resulting rHMPV-SHs virus was readily recovered and amplified. It was indistinguishable from wild-type rHMPV or biological CAN97-83 with respect to multistep growth kinetics in vitro (data not shown). To date, an analysis of the SHs gene contained in 12 independent preparations of rHMPV-SHs or derivatives and involving up to six
passages in vitro demonstrated an absence of detectable frameshifts or other mutations within the region from the end of M to the beginning of L (data not shown).

The replication efficacy of rHMPV-SHs in Golden Syrian hamsters was evaluated following intranasal administration (6, 24), with animals sacrificed on days 3 and 5 for virus titration of lungs and nasal turbinates (Table 2). There were no significant differences ($P \leq 0.05$) in the virus titers in the nasal turbinates or lungs for rHMPV-SHs compared to those for the rHMPV parent, indicating that there were no significant differences in the efficiency of replication between these two viruses.

We also evaluated the rHMPV-SHs in the AGM, an experimental animal that is more closely related anatomically and phylogenetically to the natural human host. AGMs were inoculated with rHMPV or rHMPV-SHs by the intranasal and intratracheal routes. Nasopharyngeal swabs and tracheal lavage samples were collected at intervals over 12 days postinfection to monitor virus replication in the upper and lower respiratory tracts, respectively (Fig. 4). This monitoring showed

### TABLE 2. Level of replication of rHMPV-SHs compared to rHMPV in the upper and lower respiratory tracts of hamsters

<table>
<thead>
<tr>
<th>Virus</th>
<th>Day of harvest</th>
<th>Nasal turbinates</th>
<th>Lungs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% with virus</td>
<td>Mean titer ($\log_{10}$ PFU/g of tissue) ± SE</td>
</tr>
<tr>
<td>rHMPV</td>
<td>3</td>
<td>100</td>
<td>5.0 ± 0.3</td>
</tr>
<tr>
<td>rHMPV-SHs</td>
<td>3</td>
<td>100</td>
<td>5.4 ± 0.2</td>
</tr>
<tr>
<td>rHMPV</td>
<td>5</td>
<td>100</td>
<td>5.8 ± 0.2</td>
</tr>
<tr>
<td>rHMPV-SHs</td>
<td>5</td>
<td>100</td>
<td>6.5 ± 0.1</td>
</tr>
</tbody>
</table>

$^a$ Hamsters in groups of 12 were administered 5.7 $\log_{10}$ PFU of the indicated virus intranasally under light methoflurane anesthesia on day 0.

$^b$ Six animals from each group were sacrificed on days 3 and 5. Nasal turbinates and lungs were harvested, and virus titers were determined.
that rHMPV-SHs and rHMPV were essentially equivalent in their kinetics of replication, mean peak titers (3.7 ± 0.1 versus 3.9 ± 0.1 in the upper and 5.7 ± 0.2 versus 5.2 ± 0.2 in the lower respiratory tracts, respectively), and durations of shedding (Fig. 4). As expected, neither virus was associated with clinical disease signs in AGMs. In addition, rHMPV-SHs and rHMPV were indistinguishable with regard to the titers of serum HMPV-neutralizing antibodies detected in these animals on day 28 following infection (P < 0.05) (Table 3). When challenged with rHMPV, animals that had been previously immunized with rHMPV-SHs or rHMPV were protected against HMPV challenge, as assayed by virus titration of nasopharyngeal and tracheal lavage samples (Table 3). Also, there were no significant differences in the titers of HMPV-neutralizing antibodies in sera collected 28 days postchallenge (Table 3). Thus, the SHs gene had no apparent effect on virus replication in vitro or in vivo, on immunogenicity, or on protective efficacy.

The frameshift mutation at position 78 of the SH gene ablates incorporation of SH protein into virus particles. The SH ORF initiates at the first ATG in the gene sequence, which is located at nucleotides 14 to 16. The A78 insertion, the most common mutation observed in SH, creates a frameshift that fuses the first 21 codons of the SH ORF to the −1 reading frame, which is open for seven additional codons. The second ATG in the SH gene occurs at positions 27 to 29 and opens a short ORF in the +1 reading frame that normally has a brief overlap with the SH ORF and terminates at positions 78 to 80 (data not shown). However, the frameshift that results from the insertion of a single A at position 78 would fuse this short upstream ORF in the +1 reading frame to the main body of the SH ORF. This ORF would encode a protein in which the first 19 amino acids of SH are replaced by 15 amino acids from the +1 reading frame, with the remainder of the SH protein intact.

In order to investigate whether this putative alternative form of SH might be expressed, we engineered the stabilized SH gene to contain the A78 insertion (SHs+A) and prepared a recombinant virus containing this SHs+A gene in place of the wild-type SH gene. The virus was amplified, and its sequence was confirmed from the end of the M2 gene to the start of the L gene. We purified rHMPV-SHs, rHMPV-SHs+A, and wild-type rHMPV by sucrose gradient centrifugation and analyzed equal amounts of protein on a 4 to 20% polyacrylamide gel followed by Coomassie blue staining (Fig. 5). Overall, the

![Figure 4. Kinetics of replication of rHMPV and rHMPV-SHs (bearing a stabilized version of the SH gene) in the upper and lower respiratory tracts of African green monkeys. Six animals per group were inoculated by the combined intranasal and intratracheal routes by using a 1-ml inoculum per site containing 10⁶ PFU of the indicated virus on day 0. Two additional animals were mock infected in parallel (control). Nasopharyngeal swab (A) and tracheal lavage (B) specimens were taken on the indicated days, and the titers of shed virus were quantified by plaque assay and are expressed as log₁₀ PFU/ml. The detection limit was 0.7 log₁₀ PFU/ml. Error bars indicate standard deviations.](image)

### TABLE 3. Immunogenicity and protective efficiency of rHMPV-SHs and rHMPV in AGMs

<table>
<thead>
<tr>
<th>Immunizing virus</th>
<th>Mean serum neutralizing antibody titer (log₁₀) ± SE at a</th>
<th>Challenge virus replication postchallenge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preimmunization</td>
<td>Nasopharyngeal swab</td>
</tr>
<tr>
<td></td>
<td>28 Days postimmunization and statistical grouping *</td>
<td>Mean duration of shedding (days) ± SE</td>
</tr>
<tr>
<td>Mock</td>
<td>&lt;1.5</td>
<td>8.0 ± 1.0</td>
</tr>
<tr>
<td>rHMPV</td>
<td>&lt;1.5</td>
<td>3.3 ± 1.2</td>
</tr>
<tr>
<td>rHMPV-SHs</td>
<td>&lt;1.5</td>
<td>0.3 ± 0.2</td>
</tr>
</tbody>
</table>

* AGMs in groups of six, except for the control (mock) group, which contained two animals, were immunized by intranasal and intratracheal infections using a 1-ml inoculum per site containing 6.0 log₁₀ PFU of the indicated virus or L15 medium (mock). The replication of these viruses following this inoculation is summarized in Fig. 4.

* Sera were collected on days 0 and 28 following the first infection, and the neutralizing antibody titer against HMPV was determined. The preinfection anti-HMPV serum titers were <1.5 (reciprocal log₂) for all animals in the study.

* Mean serum neutralizing antibody titers (28 days postimmunization) were assigned to statistically similar groups by analysis of variance and the Tukey-Kramer post hoc test. Values within a column that share a common letter are not significantly different (P < 0.05).

* On day 28, AGMs from each group were challenged intranasally and intratracheally with 6.0 log₁₀ PFU of rHMPV. The level of virus replication is expressed as the geometric mean of the peak virus titers (log₁₀ PFU/ml) ± the standard error for the animals in each group irrespective of sampling day. The lower limit of detection is 0.7 log₁₀ PFU/ml. A value of <0.7 log₁₀ PFU/ml is assigned to samples with no detectable virus.
pattern of viral proteins was very similar for each virus, indicating the absence of a dramatic change in virion composition or yield, as expected. Since the SH protein is undetectable by Coomassie blue staining (6), its presence in virions was investigated by Western blot analysis with a rabbit antiserum raised against a peptide containing SH amino acids 82 to 96, which represent part of the SH ectodomain (B). The HMPV proteins are indicated on the right. SHg1 and SHg2 represent two populations of SH protein that differ by the extent of glycosylation. Molecular markers are shown, with molecular masses in kilodaltons indicated at the left.

**DISCUSSION**

During the passage of HMPV in two established primate cell lines, viral subpopulations frequently emerged that contained various insertions, deletions, and point mutations. Remarkably, 31 of the 40 recombinant virus preparations that were examined contained readily apparent mutant subpopulations. A high incidence of mutant subpopulations was noted for both recombinant and biologically derived preparations. The mutant subpopulations were confirmed by sequence analysis of cDNA clones and by biological cloning.

This high incidence of detectable mutant subpopulations was surprising. It is well known that RNA viruses sustain mutations at a frequency of approximately $10^{-3}$, which is approximately equivalent to one mutation per HMPV genome, and thus exist as quasispecies. This provides the capacity for the rapid outgrowth of variants in response to selective pressure. For example, the propagation of HRSV in the presence of F protein-specific neutralizing antibodies readily yielded antibody-resistant mutants (1). However, in the absence of such pressure, the consensus sequence of a viral population can be stable during passage in vitro. For example, a comparison of the consensus sequences of two preparations of HRSV strain A2, one that was maintained in the freezer since the mid-1960s and a derivative that had been passed and plaque purified extensively in the laboratory during an intervening period of three decades, revealed only 15 nucleotide differences involving eight amino acid differences (12). Both isolates exhibited a wild-type phenotype in chimpanzees, suggesting that the changes were inconsequential, consistent with stability in vitro. In contrast, the HMPV mutant subpopulations were detected frequently and, since the recombinant viruses were made from correct cDNA clones, emerged within a few passages.

Most of the HMPV mutations involved the SH gene. Indeed, each of the 31 virus preparations contained a subpopulation with at least one frameshift (usually an insertion) that interrupted the SH ORF. In 24 preparations, the frameshift occurred at or before position 78 and would thus delete most of the molecule, including the signal/anchor sequence located at amino acids 36 to 43 (Fig. 1). Other preparations had frameshifts at position 210, 300, 397, or 407, which would result in products containing the N-terminal 65, 95, 127, or 131 amino acids of SH fused to 1, 30, 3, or 1 amino acid contributed by the alternative frame downstream of the frameshift (Table 1 and Fig. 1). This mutation would truncate the C-terminal extracellular portion by 114 to 48 amino acids but would leave the transmembrane and cytoplasmic coding regions intact. However, drastic mutations to a glycoprotein often disable it for folding and intracellular processing and it is unknown whether these truncated forms of SH were stable. In any event, there was a consistent pattern of ablating the expression of full-length SH protein.

The functions of the SH protein of HMPV, like that of HRSV, are unknown. The HRSV SH protein appears to assemble into homopentamers, and a channel-like activity is suggested by an increased membrane permeability to low-mass compounds observed when the SH protein is expressed in bacteria (11, 17, 20). Studies with simian virus 5 and mumps virus suggest that SH antagonizes the induction of apoptosis by tumor necrosis factor alpha and might also block the activation of NF-κB (29). Although SH appears to be completely dispensable for replication in vitro and is nearly so in AGMs (see the introduction), every direct clinical isolate of HMPV described to date contains an intact SH gene and SH ORF (16, 25). Thus, the emergence of mutants in which SH has been silenced appears to be specific to cell culture.

The fact that the SH gene is nonessential would allow mutations to readily accumulate without compromising replication. However, the dispensability of SH seems insufficient to account for its high incidence of mutation because G is equally dispensable for growth in vitro and M2-1 and M2-2 are nearly so (6, 7). The high incidence of mutations also did not seem to be due to unusual features of the SH gene sequence; for example, the 627-nucleotide SH gene of this strain contains 16 tracts of four or more A or T residues, approximately 75% more than the 10 such tracts contained in the 711-nucleotide G
gene. However, no insertions were observed in G compared to the 41 insertions (and two deletions) observed in SH. The insertions in SH occurred most frequently (21 instances) in the run of seven A residues at position 78, making that somewhat of a “hot spot.” However, there were 22 other instances of insertions in SH involving 12 other sites (and two instances of deletions involving two sites).

The observation that each of the 31 mutant populations contained a frameshift in the SH ORF suggested an alternative explanation, namely, that ablating the expression of the full-length SH protein provided a selective advantage in cell culture. This proposed advantage presumably is not very potent because otherwise there would have been a more rapid and complete dominance of virus with the silenced SH gene. The idea that an adaptive mutation might occur without an obvious advantage in growth is not without precedent. For example, a clinical isolate of human parainfluenza virus type 2 accumulated a T-to-C nucleotide substitution at nucleotide 15 in the leader region during passage in vitro that appeared to represent a subtle adaptation to cell culture, although it did not confer a detectable growth advantage and was attenuating in AGMs (S. M. Nolan et al., unpublished data). The present example with this clinical isolate of HMPV is somewhat unusual in that it involves the silencing of an entire gene rather than one or a few point mutations.

We suggest that each mutant subpopulation arose beginning with a frameshift in the SH ORF. This frameshift conferred a modest selective advantage that allowed the mutant to expand into a detectable subpopulation. In the process of expanding, the silenced SH gene would be a blank slate for accepting and retaining whatever mutations occur during replication in vitro. The idea that a frameshift was the first event in a stepwise acquisition of mutations was suggested from the analysis of cDNA clones, which showed that some molecules in a subpopulation contained only a frameshift, whereas others contained this frameshift combined with additional mutations. We suggest that the panel of observed mutations (Table 1) provides a glimpse of the typical types of mutations that HMPV sustains during replication in vitro; ordinarily, these are not borne by a preferentially amplified molecule and, thus, are not easily detectable. By this measure, small insertions are the most frequent mutational event, with point mutations and biased hypermutations being somewhat less frequent and deletions being the least frequent. In most regions of the genome, a frameshift would be detrimental to viability; thus, if this indeed is the most common mutation that occurs during HMPV replication, it would contribute to a population of defective molecules. Unlike point mutations, for the most part, these frameshifted genomes would not contribute to a viable quasispecies but rather seem to be a dead end.

Inspection of the individual insertion sites did not reveal any common motif upstream or downstream of the sites; the only common feature seemed to be the homo-oligomer tract. The insertions thus likely involve the familiar mechanism of polymerase stuttering (14). Our limited sampling suggests that this occurs promiscuously: this particular SH gene contains 12 runs of four or more A residues, and seven of these were used in the observed insertions. The SH gene sequence also contains four runs of four or more T residues, and two of these were used in the observed insertions. Out of a total of 46 insertion events, 42 (91%) involved A and the remainder involved T. However, much of this apparent bias towards A (mRNA sense) might be dictated by the characteristics of the viral sequence. For example, the SH gene contains threefold more runs of four or more residues involving A than those involving T and it lacks runs of four C or G residues. The rest of the HMPV genome has a similar bias.

The idea that frameshifting in the homo-oligomer tracts of SH was key to generating the mutant subpopulations was confirmed by replacing the naturally occurring SH gene with a synthetic version, in which these tracts had been silently ablated. This greatly reduced the incidence of detectable mutations in the recombinant viruses in which it has been used (data not shown). We anticipate that adventitious mutations at nonessential sites will continue to occur but usually will not be amplified into significant subpopulations. In any event, virus preparations for vaccine purposes should always be sequenced in entirety and such occurrences would thus be detected.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the NIAID, NIH.

We thank Jens Fricke and Cindy Luongo for helpful discussions, Quynh Pham and Emerito Amaro-Carambot for excellent technical assistance, and Guy Boivin (Laval University, Quebec, Canada), Teresa Peret, Dean Erdman, and Larry Anderson (CDC, Atlanta, Georgia) for providing an initial inoculum of HMPV CAN97-83.

REFERENCES


