JOURNAL OF VIROLOGY, Dec. 1995, p. 7768–7774 Vol. 69, No. 12
0022-538X/95/$04.00+0

Evolution of the Sindbis Virus Subgenomic mRNA Promoter in Cultured Cells

JENNIFER M. HERTZ AND HENRY V. HUANG*

Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110-1093

Received 27 April 1995/Accepted 30 August 1995

Transcription of the subgenomic mRNA of alphaviruses initiates at an internal site, called the promoter, which is highly conserved. To determine the functional significance of this conservation, we used an approach that randomizes positions −13 to −9 of the promoter to generate a library containing all possible sequences within this region, including the wild-type sequence. Viruses in the mixed population with more-efficient promoters were selected for during passaging in mammalian (BHK-21) cells. Results from early passage populations indicate that a large number of different promoters are functionally active. Analysis of eight individual viruses found that although each contained a promoter with different degrees of sequence identity to the wild-type sequence, all eight viruses produced progeny. This suggests that the mechanism for transcription allows for a diversity of sequences to serve as promoters. Further passaging of the viral library led to a population consensus sequence that increasingly resembled the wild-type sequence, despite the fact that these promoters are not constrained by the need to encode the carboxyl terminus of the nsP4 protein. Thus, conservation of the region of the promoter from −13 to −9 is in large part due to selection for promoter function, and the wild-type sequence and sequences closely similar to it seem to be optimal for promoter function in BHK-21 cells.

Sindbis virus (SIN) is an enveloped positive-strand RNA virus in the *Alphavirus* genus of the family *Togaviridae* and infects mosquito, avian, and mammalian species. In an infected cell, the SIN genome is translated to produce two polyproteins that are ultimately cleaved to produce four nonstructural proteins, nsP1 through nsP4 (3). These proteins are necessary for viral replication, which is initiated by using the genomic RNA as a template for minus-strand synthesis (for reviews, see references 20 and 21). The minus strand is then used as a template for synthesis of daughter full-length plus-strand RNAs. In addition, the minus strand is used as a template for transcription of a smaller or subgenomic mRNA. Transcription utilizes an internal promoter to produce these subgenomic mRNAs (4, 8), which are identical to the 3′ third of the genomic RNA. The subgenomic mRNAs encode the viral structural proteins (15), which are required for progeny virion production.

A comparison of known alphavirus genomic sequences identified a 21-nucleotide conserved sequence called the junction region two thirds from the 5′ end of the genomic RNA (11). This region codes for the carboxyl terminus of nsP4 on the plus strand and also contains the initiation site for subgenomic mRNA synthesis on the minus strand. Deletion mapping showed that the minimum nucleotide sequence necessary for transcription of the subgenomic mRNA extends from nucleotide −19 to nucleotide +5 relative to the subgenomic mRNA initiation site (4, 8). This minimal promoter encompasses the conserved junction sequence (11) and is active in both defective interfering genomes and in SIN, although it is less active than, e.g., a promoter from −98 to +14 (8, 12).

The sequence conservation in the promoter region could be the result of chance, of nsP4 amino acid coding constraints, or of nucleotide coding constraints for promoter recognition. The expected mutation rate of viral RNA polymerases is approximately 1 nucleotide per 10^4 to 10^5 nucleotides synthesized (17, 19). This high mutation rate makes it unlikely that the nucleotide conservation is due to just chance. For alphaviruses, the sequence coding for nsP4 terminates at different positions in the promoter region, yet sequences beyond the termination codon are conserved. The wobble positions of the nsP4 codons are expected to be mutable in the absence of other selection, yet they are well conserved. This suggests that the sequence conservation might not be primarily constrained by the amino acid sequence of nsP4 but is due to selection for promoter function.

In support of this, we found that junction region sequences from other alphaviruses that correspond to the SIN minimal promoter function as promoters for SIN (4). So, although base changes within the minimal promoter have arisen with speciation, promoter function has been conserved. Are there other sequences that can function as promoters? Can base changes at the absolutely conserved nucleotide positions be tolerated? If they can, will they function as well as or better than the wild-type sequence, or has evolution presented us with the optimal promoter? In general, what are the sequence requirements for promoter function? It is unknown whether the primary sequence is critical or whether other influences such as RNA structure and sequence context play a role in the recognition of the promoter. Classical mutagenesis results showed that single-base changes can alter promoter activity (13). However, a thorough analysis of the promoter region by this approach would require the construction and analysis of a large number of individual viruses containing different promoters with single, double, triple, etc., mutations and combinations thereof, and this would be impractical. Instead, a more efficient approach that selects for active promoters among a randomized library was used.

* Corresponding author. Mailing address: Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, 660 S. Euclid Ave., St. Louis, MO 63110-1093. Phone: (314) 362-2755. Fax: (314) 362-1232.

MATERIALS AND METHODS

General recombinant DNA materials and methods. Restriction enzymes and SP6 DNA-dependent RNA polymerase were obtained from New England BioLabs, Inc., Bethesda Research Laboratories, Epicentre Technologies, or Boehringer Mannheim Biochemicals. Radioactive materials were obtained from Amersham Corp. or ICN Pharmaceuticals, Inc. Enzymes were used essentially as recommended by the manufacturers. Plasmids were grown, purified, and analyzed by standard methods, with minor modifications (10).

cDNA library of SIN derivatives. A library of SIN derivatives was constructed in cDNA plasmids (see Fig. 1). The virus genomes have two promoters, one containing the wild-type sequence and the other containing a randomized region of 5 nucleotides. The mutated promoters were generated by PCR with the SIN cDNA plasmid Toto1000 as a template (14). One of the primers contains a randomized region and hybridizes to nucleotides 7611 to 7569 (SIN coordinates). The randomized nucleotides are from 7585 to 7589, corresponding to positions −13 to −9 of the promoter, where +1 is the start site of the subgenomic mRNA. This 3′ primer also contains an *Xba*I site for cloning, and the 5′ primer contains *Xho*I and *Apa*I sites adjacent to SIN nucleotides 7500 to 7521. The amplified fragment of 136 bp was purified, digested with *Xho*I and *Xba*I, and cloned into a shuttle vector (PneoS) digested with the same restriction enzymes. This placed the amplified fragment which contained full-length promoters (from −98 to +14) directly upstream of the coding region for the structural proteins.

The subcloned library was digested with *Xho*I and *Bss*HII (SIN 9804) and ligated with a deletion vector called TDV digested with the same restriction enzymes to produce a cDNA library of full-length viruses (see Fig. 1). The 5′ portion of each virus consists of wild-type SIN sequences up through the wild-type promoter. Downstream of this first wild-type promoter is the bacterial chloramphenicol acetyltransferase (CAT) coding sequence, followed by the randomized second promoter, which is upstream of the coding region for the structural proteins. Hence, each virus contains two promoters and can potentially produce two subgenomic mRNAs. The first subgenomic mRNA is transcribed from the wild-type or CAT promoter (upstream of the CAT coding region) and is called the CAT mRNA. The second subgenomic mRNA transcribed from the structural or STR promoter, containing the randomized region, is called the STR mRNA. Since the randomized region is 5 nucleotides long, the maximum potential number of different cDNAs in the library is 4^5, or 1024. Twenty-two of the randomized sequences with the addition of flanking nucleotides can constitute *Xba*I or *Xho*I sites. Since these restriction enzymes were used to construct the library, these sequences may be underrepresented or absent. The full-length cDNA library constructed contains 3,400 clones, so each species should be represented three to four times. A cDNA plasmid of a virus with a wild-type STR promoter, called TCS (short for Toto1000, CAT, and STR), was also constructed by the same methods except that the 3′ primer contained no randomized sequences, so both its promoters are wild-type.

Construction of virus 19/5. The genome of virus 19/5 is identical to that of TCS, except that the STR promoter contains the minimal 24-nucleotide promoter sequence, from −19 to +5, instead of the full-length promoter. The cDNA of this virus was constructed as above, but the minimal promoter (−19 to +5) was derived from JSINCAT (4). JSINCAT and PneoS were digested with *Xba*I and *Pvu*I. The appropriate bands were isolated from a 1% low-melting-temperature gel and ligated to produce 19/5Str. This clone contains the minimal promoter upstream of the entire structural protein coding sequence. 19/5Str was digested with *Asp*718, and the overhang was filled with Pol I Klenow and then digested with *Bss*HII. As described above, TDV was digested with *Xho*I, but then the ends were also filled and then digested with *Bss*HII. The appropriate bands were isolated and ligated to produce the 19/5 cDNA containing the full-length virus sequence, having a 24-nucleotide STR promoter.

RNA transcription and transfection. In vitro synthesis of 5′-capped RNA transcripts was at 40°C for 45 min with SP6 DNA-dependent RNA polymerase from DNA samples linearized with *Sst*I (2). The samples were then treated with DNase I (Bethesda Research Laboratories) for 30 min at 37°C to remove the template. RNA (5 μg) was transfected by electroporation (9) into 2×10^7 baby hamster kidney BHK-21 (C-13) cells (ATCC CCL10) grown at 37°C in minimal essential medium with Earle's salts (MEM), supplemented with 10% heat-inactivated fetal calf serum.

Infectious center assays. Infectious center assays were performed to determine the number of electroporated cells that can produce plaque-forming viruses. This assay provides a minimum estimate of the total number of infectious RNA species successfully electroporated into the cells. One-fiftieth of the electroporated cells (4×10^5) was serially diluted for plating. After the cells were allowed to attach to the plates, 10^6 cells were added to each plate to form a complete monolayer. At 4 h postelectroporation, a 1:1 mixture of 1.8% agarose and plaquing medium (2× minimal essential medium with Earle's salts, 6% heat-inactivated fetal calf serum, 200 U of penicillin per ml, and 200 μg of streptomycin per ml) was used to overlay the monolayers. The plates were incubated for 2 days at 37°C. Plaques were detected by staining with crystal violet.

Passaging the viral library in BHK-21 cells. Electroporated cells not used for infectious center assays were cultured, and the medium was harvested at 15 h postinfection (hpi) at 37°C, resulting in a total yield of 5×10^7 viruses. The medium at this first passage is called the P1 viral population, while the RNA isolated from the electroporated cells, P0 RNA, represents the initial library. The P1 medium was used to infect BHK-21 cells to harvest the P2 medium and the RNA that represents the P1 viral population (P1 RNA), at 5 hpi. Medium and RNA were collected and harvested through passage 4. A multiplicity of infection (MOI) of 0.1 was used for each passage, with 2×10^5 PFU to infect 2×10^6 cells.

Reverse transcription PCR. Total RNA was isolated from cells with RNAzol B (Cinna/Biotecx). In a 20-μl reaction mixture, 900 ng of total RNA was reverse transcribed with Superscript II (Bethesda Research Laboratories) and a 3′ primer that hybridizes to nucleotides 7990 to 8007 of SIN, within the capsid coding region. A 10-μl sample of the reaction mix plus a 5′ primer within the CAT coding region and 2.5 U of *Tth* DNA polymerase (Epicentre) per 100 μl of reaction mixture was used to amplify a 682-bp fragment containing the STR promoter.

Cycle sequencing. The specifically amplified fragment was purified from 1/5 to 1/10 of the reverse transcription PCR products resolved through a 2% Nusieve gel (FMC BioProducts). To sequence the purified DNA, the ΔTAQ kit (U.S. Biochemicals) was used according to the manufacturer's protocol for kinased primers. The primer used hybridizes from nucleotide 7671 to 7649 of SIN at the 5′ end of the capsid coding sequence, about 60 nucleotides from the randomized region. The samples were electrophoresed through an 8% polyacrylamide sequencing gel and then exposed to film at −20°C.

Subcloning amplified promoter regions. The amplified fragments of 682 bp obtained from reverse transcription PCR were isolated from a 2% Nusieve gel, digested with *Aat*II and *Xho*I, reisolated after electrophoresis as a 522-bp fragment through a 2% Nusieve gel, and subcloned into PneoS isolated after digestion with the same restriction enzymes. These clones were used for sequence analysis of individual promoters within each population and for construction of full-length viruses, as described below.

Construction of viruses with promoters selected from the library. Subcloned STR promoters were digested with *Xho*I (just upstream of the STR promoter region) and *Bss*HII (SIN 9804) and isolated from a 2% Nusieve gel. Each fragment was ligated to TDV digested with the same restriction enzymes to produce cDNA clones containing the full-length viral sequence.

RNA analyses. BHK-21 cells were infected by each virus at an MOI of 5. Dactinomycin (1 μg/ml) was added 1 h before labeling. The cells were incubated with 100 μCi of [^{3}H]uridine per ml for 2 h, at which time (5 and 9 hpi) intracellular RNA was isolated with RNAzol B. The RNA was denatured with glyoxal and dimethyl sulfoxide and electrophoresed through 1% agarose gels (1). The gels were treated for fluorography (6) and exposed to film from 2 days to 1 week, and the autoradiographs were used as guides to cut out the two subgenomic mRNA bands in each sample for quantitation by liquid scintillation counting. The molar ratio of STR mRNA to CAT mRNA was then calculated by dividing the counts in the STR mRNA band by the counts in the CAT mRNA band and then multiplying the result by 1.2 to correct for the mRNA size difference.

Statistical analyses of individual promoters from viral populations. Within a population, the actual and expected numbers of promoters sequences that have zero to five matches to the wild-type SIN promoter sequence were compared by the chi-square test (18). The contingency chi-square analysis (16) was used to compare promoter sequences from different populations. Significance was assigned at the level of $P < 0.05$.

RESULTS

Experimental approach for selection of functional promoters. A library of mutagenized SIN viruses was constructed to determine which nucleotides are important for promoter recognition and subsequent transcription of the viral subgenomic mRNA. A viral RNA library was created with a modified SIN with two promoters (Fig. 1). The promoter that directs the transcription of the structural protein mRNA extends from nucleotide −98 to +14 relative to the start site of the mRNA. The nucleotides at positions −13 to −9 of this promoter are randomized. Therefore, the maximum number of viruses in the population is 4^5, or 1024. The activity of these promoters determines the amount of structural proteins available for progeny virion packaging. Therefore, viruses that have promoters most consistent with the viral life cycle are expected to produce more progeny virions and are selected for by passaging the viral populations at a low multiplicity of infection (MOI). For example, given hypothetical viruses *W* and *L*, if virus *W* has a certain promoter activity and produces 10^4 progeny virions per infected cell while virus *L* has a relatively lower level of promoter activity and produces only 10^3 progeny virions per infected cell and if the starting numbers of each virus are equal, then the medium from passage 1 would contain

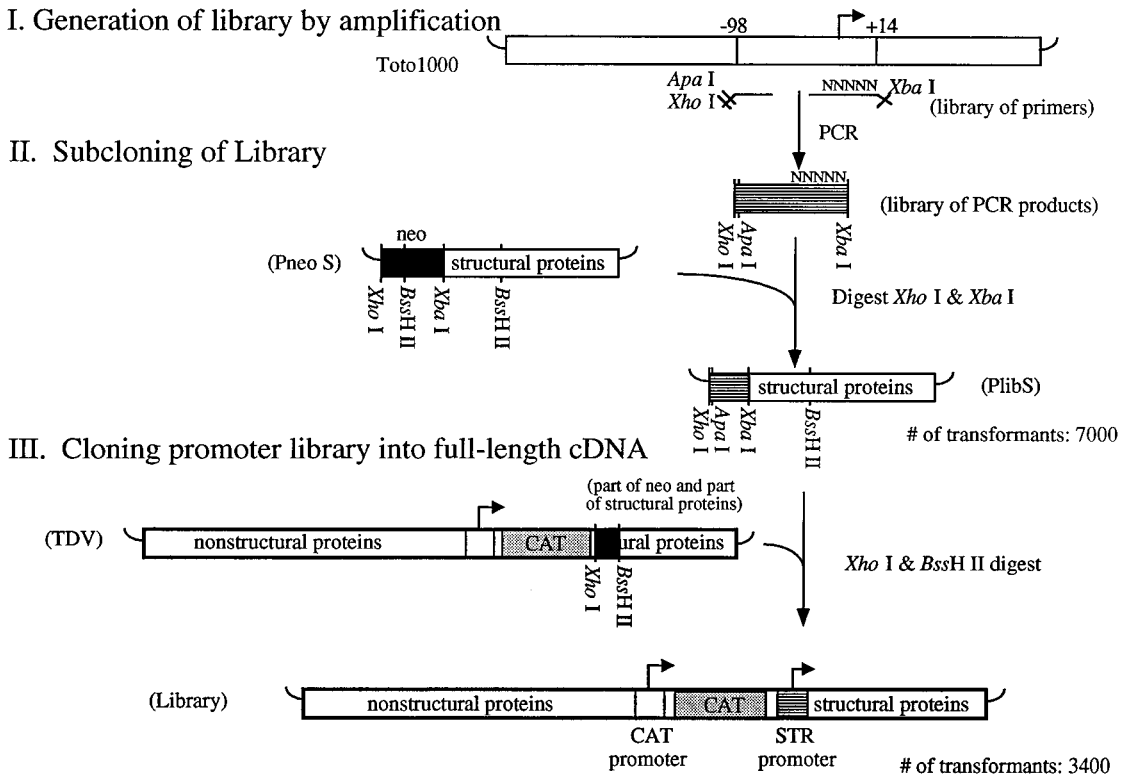

FIG. 1. Construction of a cDNA library of SIN derivatives. The SIN derivatives in the library contain two promoters. The first, called the CAT promoter, is the wild type with the CAT gene downstream. The second, called the STR promoter, contains a block of 5 randomized nucleotides and has the structural protein coding region downstream. (I) The library of STR promoters was generated by PCR with an oligonucleotide containing a randomized sequence corresponding to positions −13 to −9 of the promoter. (II) The PCR products were subcloned into a vector, PneoS, upstream of the structural protein coding sequence to produce the PLibS library (7,000 transformants). (III) The PLibS library was cloned into the TDV vector to produce the library of full-length viruses. The maximum number of different cDNAs in the library is 1,024 (4^5). The final library consists of 3,400 transformants; thus, each possible sequence is represented three to four times on average.

10-fold more *W* progeny than *L* progeny. During passage 2, there would be 10 times more *W*-infected cells, each of which produces 10 times more progeny virions than the *L*-infected cell. Thus, the *W* virus would become 100 times more abundant in the passage 2 medium, and the *L* virus would be only 1% of the viral population. Passaging at a low MOI minimizes coinfections that might allow interactions between viruses within an infected cell, whereby a virus with a higher level of promoter activity might be able to complement the defect of another virus with a lower level of promoter activity. It should be noted that the virus with the most active promoter might not be biologically optimal. If overexpression of the viral mRNA occurred at the expense of, e.g., genomic RNA replication, the number of genomic RNAs available for packaging would be reduced. Thus, passaging is expected to select for promoters with activities consistent with the entire viral life cycle.

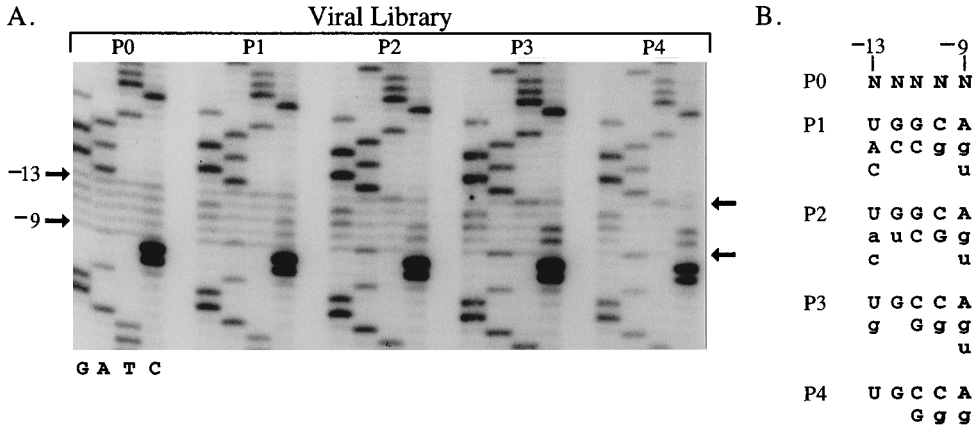

FIG. 2. Passaging selects for specific promoter sequences. (A) The viral library was passaged in BHK-21 cells. The RNA from passage 0 through passage 4 was isolated and reverse transcribed, and the STR promoters were PCR amplified. The sequence of the promoters in each passaged population is shown. (B) The RNA consensus sequence of the populations at each passage is derived from those bases that are most abundant at each nucleotide position.

TABLE 1. Promoter sequences from passaging in BHK-21 cells at 37°C

No. of nucleotide matches with 3′UGCCA	Sequence (virus)[a]	Frequency of sequence in[b]:		
		P0	P2	P4
5	UGCCA (II)		2	1
4	AGCCA		1	1
4	UACCA		1	
4	UGCCU		1	
4	UGGCA		2	3
4	UUCCA		1	
3	AGACA		1	
3	AGCCU		1	
3	CGGCA		2	
3	GGCCG		1	
3	UAUCA		1	
3	UACCG (A1)		1	
3	UGACG	1		1
3	UGCGG		1	
3	UGGAA		1	
3	UGGCU			3
3	UGGCG		1	1
3	UGGGA (III)	1	4	2
3	UUCGA		1	
3	UUCUA		1	
2	AGCGU		1	
2	AGCUG		1	
2	AGGCG		1	
2	AGGGA		1	
2	AUACA		1	
2	CGCUU (B2)		1	
2	CGUCG		1	
2	GUACA		1	
2	GUGCA			1
2	UCACG	1		
2	UCGCG	1		
2	UGAUG		1	
2	UGGGG		3	
2	UGGGU		2	
2	UGUAG		1	
2	UGUGU		2	
2	UGUUG			1
1	AAACC (C3)		1	
1	AGGGG		1	
1	AGUUG			1
1	AUUCG		1	
1	CGUUU		1	
1	GUCUU		1	
1	GUGGA	1		
1	UAGAU			1
1	UCAGU		1	
1	UUAGG	1		
0	AAGUG		1	
0	AUAAU	1		
0	AUAGU	1		
0	GCUUG	1		
0	GUAAG	1	1	
0	GUUGG		1	

[a] The sequences from −13 to −9 (3′ to 5′) of the STR promoter from individual clones isolated after passaging in BHK-21 cells at 37°C are shown. The STR promoters for II, A1, III, B2, and C3 were cloned into full-length viruses for further analysis.

[b] P0, P2, and P4 represent the viral populations at passages 0, 2, and 4, respectively. Shown under each column is the number of times the sequence was found in that passage.

Selection by passaging viruses in tissue culture. A mixed population of two viruses was used to verify that passaging at a low MOI selects for viruses with functional promoters. cDNA from a previously described low-titer SIN mutant (2), Toto1100CR4.1, that contains a 3-nucleotide insertion within the promoter was mixed at a ratio of 1,000:1 with the wild-type Toto1100. The cDNA mix was transcribed in vitro, and the RNA transcripts were transfected into BHK-21 cells. Medium from the first passage that contains the P1 viral population was collected at 15 h posttransfection, and the intracellular P0 RNA was isolated. The P1 viral population was used to infect BHK-21 cells at an MOI of 0.1, and the intracellular P1 RNA was isolated 5 hpi. P0 and P1 RNAs were reverse transcribed and the promoter region was PCR amplified for sequence analysis.

Cycle sequencing of the promoter region was used to estimate the relative abundance of each virus in the P0 and P1 populations. Using mixtures of defined molar ratios of mutant versus wild-type RNA transcripts of the respective cDNA clones, we found that the detection of a specific sequence by cycle sequencing requires that it constitute ≥20% of the total RNA (data not shown), consistent with other estimates (7). As expected with the input 1,000:1 ratio of mutant-to-wild-type RNA transcripts, cycle sequencing of the P0 RNA population (from the transfected cells) detected only the mutant promoter (data not shown), indicating that at 15 hpi there is little or no progeny from potential reinfections. In contrast, the sequence representing the mutant promoter was no longer detectable in the P1 RNA (data not shown). Therefore, the wild-type virus was enriched at least 5,000-fold after a single passage. Thus, passaging at a low MOI does provide strong selection for functional promoters.

Construction of a viral library to select for functional promoters. The alphavirus promoter normally overlaps sequences encoding the nsP4 protein. To restrict the base changes to the promoter alone, the SIN genome was modified to contain two promoters (Fig. 1). The first, called the CAT promoter, is a wild type and is at the normal position in the SIN genome. It is immediately upstream of the CAT gene, and it directs the transcription of the CAT mRNAs. The second, called the STR promoter, is located downstream of the CAT gene and directs the transcription of the STR mRNAs that encode the structural proteins. Positions −13 to −9 of the STR promoter were randomized to create a library of viruses consisting of 1,024 possible STR promoter sequences. The CAT sequence was placed between the two promoters to minimize the potential influence of one promoter on the other (12). It also places the structural protein sequences downstream of the CAT sequence in the CAT mRNA, thus minimizing any potential translation

TABLE 2. Consensus of individual STR promoters from viral populations

Passage and base	% at position[a]				
	−13	−12	−11	−10	−9
2					
A	24	10	14	6	**42**
C	10	2	**32**	**44**	2
G	10	**70**	**38**	**36**	**34**
U	**56**	18	16	14	22
4					
A	12	6	6	6	**50**
C	0	0	12	**69**	0
G	6	**88**	**69**	12	25
U	**81**	6	12	12	25

[a] Shown is the percentage of each base at positions −13 to −9 of the STR promoter on the minus-strand RNA, for passages 2 and 4 (data from Table 1). Values in bold type (most abundant at each position) make up the RNA consensus sequence for each passage. The wild-type sequence is 3′UGCCA.

of the structural proteins from this bicistronic mRNA. The CAT promoter conveniently serves as an internal reference for normalizing STR promoter activity.

To maximize the probability of recovery of any of the 1,024 sequences within the library that are capable of functioning as promoters, the efficiency at each step of library construction was monitored to ensure adequate statistical sampling. The steps that have the greatest potential for random loss of individual sequences are (i) construction of the cDNA libraries and (ii) transfection of host cells. Because of the constraints of transformation and transfection efficiencies, the length of the randomized region was limited to 5 nucleotides to ensure that all possible sequence combinations are represented multiple times in the library at all stages. The library used here was derived from 3,400 independent full-length cDNA clones, so that each species is represented three to four times on average. The transfection efficiencies of electroporation, conservatively measured by the number of infective centers, ranged from 2×10^4 to 5×10^4, so each species should be represented at least 20 to 50 times in the transfected cells. Overall, there is a 95% probability, on the basis of the Poisson distribution, that all 1,024 species are present in the library and were electroporated into the cells.

Promoters selected for by passaging in BHK-21 cells. RNA transcripts of the cDNA library were electroporated into BHK-21 cells (9), and the medium was harvested after 15 h. The viruses were further passaged at an MOI of 0.1 in BHK-21 cells. Overall, plaque morphology of the population became more uniform with passaging. The total yields of the viral populations at each passage were from 10^5 to 10^6 PFU. Intracellular RNAs were isolated for analysis of the promoters present in each passaged population. The RNA from each passage was reverse transcribed, and the region containing the STR promoter was PCR amplified. Specifically amplified products were then isolated and used for cycle sequencing. Figure 2A shows the DNA sequence patterns that represent the STR promoters of viral populations isolated from P0 through P4 cells. The P0 passage RNA shows equivalent representation of each base at each of the five nucleotide positions (−13 to −9), as would be expected for a randomized sequence. With further passaging (lanes P2 through P4), specific bases at each nucleotide position became progressively enriched while others became depleted. A well-defined consensus sequence evolved by passage 4 (Fig. 2B), and it shows a high degree of homology to the wild-type SIN sequence, 3′UGCCA.

To look at the individual promoters that make up the populations, individual STR promoters were cloned and sequenced (Table 1). Ten isolates from the P0 population were sequenced, and they were all different. In 50 isolates from the P2 population, 40 different sequences were found, 7 of which, including the SIN wild-type sequence, were found two or more times. A sample of the promoters from this population was studied individually (see below). They had a range of promoter activities, but all were sufficiently active for the production of progeny virions. Sixteen isolates from the P4 population had 11 different sequences, 3 of which were repeatedly isolated (Table 1).

The consensus sequence derived from the individually sequenced promoters from each passage (Table 2) is very similar to that of the corresponding population as a whole (Fig. 2B). For example, the consensus of the 50 individual sequences from P2 resembles the sequence of the P2 population as a whole (Table 2; Fig. 2B). Thus, the sequence of the population as a whole is an accurate representation of the constituent, individual promoter sequences.

One way to measure the evolution of the population during passaging is by counting the number of individual STR promoter sequences that have zero to five base matches with the wild-type SIN sequence, 3′UGCCA. The number of the individual sequences in each class can be compared with what a random population would contain. As expected, the sequences from P0 have the same distribution as a random population ($P > 0.9$). However, the P2 and P4 sequences do not resemble the random population ($P < 0.005$); i.e., selection has occurred in a nonrandom manner during passaging. As shown graphically in Fig. 3, the P0 population is only a few fold enriched or depleted for sequences with zero to five matches to 3′UGCCA, compared with that expected of a random population. However, passages 2 and 4 gave 40- to 60-fold enrichments, respectively, for the wild-type sequence. Similarly, promoters with three or four matches with the wild-type sequence were enriched with passaging, while those with two matches or fewer were depleted. These data clearly show that the promoters in the viral population evolved during passaging from a random library to sequences that resemble the SIN wild-type promoter.

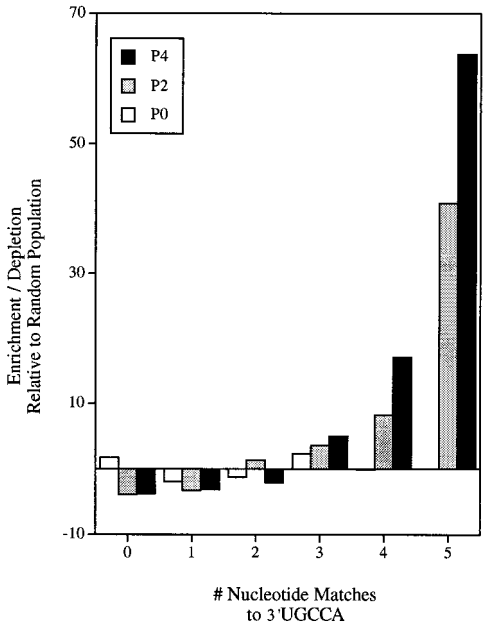

FIG. 3. Enrichment of the wild-type promoter sequence during passaging in BHK-21 cells at 37°C. For each passage, the individual STR promoters that were sequenced were divided into groups whose positions −13 to −9 have zero to five nucleotide matches with the SIN wild-type sequence, 3′UGCCA. The enrichment (positive value) of each group is defined as the actual number of promoters in the group divided by the number expected if the population had been random. The negative inverse is plotted for groups that show depletion.

Growth and RNA analysis of individual promoters selected from the viral library. Eight STR promoters were further analyzed. The promoters of viruses TCS and II have the wild-type sequence. Virus II contains an independently derived wild-type STR promoter. Three of the STR promoters analyzed were found after passaging the viral library in mosquito cells (5). To eliminate the confounding effects of potential second-site suppressors (2), the subcloned promoters were used to construct full-length viruses with double promoters (see Materials and Methods). In addition, virus 19/5, which contains an STR promoter with only the minimal promoter sequence, −19 to +5 (4, 8, 12), was constructed. The RNA transcripts for viruses A1, A3, A7, A11, B2, C3, II, III, TCS, and 19/5 were electroporated into cells to produce viral stocks.

BHK-21 cells were infected with each virus at an MOI of 5.

TABLE 3. RNA analysis and growth of individual viruses

Virus	Sequence[a]	PFU (10^7)/ml at 3 to 5 hpi (decrease [fold])[b,c]	PFU (10^8)/ml at 7 to 9 hpi (decrease [fold])[b,c]	STR/CAT molar ratio (decrease [fold]) at[c,d]:	
				5 hpi	9 hpi
TCS	UGCCA	2.5 (1)	7.0 (1)	2.1 (1)	2.5 (1)
II	-----	2.9 (1)	8.0 (1)	1.8 (1)	1.9 (1)
A3	----G	0.3 (10)	4.5 (2)	1.5 (1)	1.5 (2)
III	--GG-	0.6 (4)	3.0 (2)	1.5 (1)	2.0 (1)
A7	---AU	0.6 (4)	2.1 (3)	1.0 (2)	1.1 (2)
A1	-A--G	0.6 (4)	0.5 (14)	0.7 (3)	1.0 (3)
A11	C-GG-	0.4 (7)	2.2 (3)	1.2 (2)	1.3 (2)
B2	C--UU	0.2 (13)	0.9 (8)	0.7 (3)	0.8 (3)
C3	AAA-C	0.03 (100)	0.1 (70)	0.2 (11)	0.2 (13)
19/5	-----	1.1 (2)	0.8 (9)	0.9 (2)	1.4 (2)

[a] The sequence of positions −13 to −9 (3′ to 5′) of the STR promoter of each virus is shown. Virus II contains an independently derived wild-type STR promoter (see Table 1). Dashes represent nucleotide identity with the wild-type sequence of TCS.

[b] BHK-21 cells were infected with each virus at an MOI of 5, media were collected, and titers were determined with BHK-21 cells.

[c] Decreases are relative to TCS.

[d] BHK-21 cells were infected with each virus at an MOI of 5, and [^{3}H]uridine-labeled intracellular RNA was isolated at 5 and 9 hpi, denatured, and separated through 1% agarose (Fig. 4). The molar ratio is equal to the amount of STR mRNA (counts per minute) divided by the amount of CAT mRNA (counts per minute) and multiplied by 1.2 to correct for mRNA size difference.

To measure the rate of progeny virus production for each virus, the medium was replaced 2 h prior to harvesting at 5 and 9 hpi for the determination of titers on BHK-21 cells. All of the viruses produced progeny virions, with TCS having the highest yields. The rate of viral release for all but one of the viruses is no more than 10-fold lower than that for TCS (Table 3). C3 is the only virus with a 100-fold decrease.

The intracellular viral RNA produced by each virus was labeled with [^{3}H]uridine and isolated at 5 and 9 hpi. Figure 4 shows that all eight viruses produce a genomic and two subgenomic RNAs. The molar ratios of the STR mRNAs to the CAT mRNAs were determined to compare the activities of the STR promoters (Table 3). Similar to the growth rates, TCS has the highest molar ratio at both 5 and 9 hpi. However, the molar ratios of the other viruses are within threefold of the TCS ratio, except for that for C3, which is about 10-fold lower. Virus 19/5 has molar ratios similar to several of the individual viruses that contain much larger promoter regions (−98 to +14), albeit with a few non-wild-type nucleotides. From these data, we conclude that all the viruses contain active STR promoters and can grow to various degrees in BHK-21 cells.

DISCUSSION

To determine the sequence requirement for alphavirus promoter function, we selected for active promoters from a library of viral sequences. There was no obvious enrichment of any particular base at any of the five randomized positions in the transfected cells. This is as expected, since the sequences should correspond to the original, randomized library. Passaging selected for viruses containing promoters with increased number of matches with the wild-type sequence, especially after four passages (Fig. 3; Table 2).

Eight different promoter sequences with one to five matches with the wild-type sequence were chosen for further analysis. All eight viruses produced STR mRNAs, and as expected, all produced progeny virions (Fig. 4; Table 3). Viruses TCS and II, which both contain the wild-type STR promoter, have the highest growth rates of all the viruses tested. At the other extreme, C3, which contains only one wild-type nucleotide in the region from −13 to −9, has the lowest growth rate, producing the least amount of progeny. The promoter strength of the STR promoter of TCS is greatest and that of C3 is the lowest for the viruses analyzed, indicating a direct correlation between promoter activity and progeny virion production. Therefore, as is observed, passaging the viral population results in a gradual loss of those viruses that have less-active STR promoters, like C3, and an enrichment of those viruses with more-active STR promoters, since the latter produce more progeny virions. The superior growth of viruses with the wild-type STR promoter (TCS and II) (see Table 3) suggests that they should be enriched during passaging, at the expense of other viruses in the population. Indeed, the wild-type promoter was isolated from both P2 and P4 populations.

It is noteworthy, however, that viruses with non-wild-type promoter sequences were readily isolated, even after four passages (Table 1). All of the non-wild-type sequences tested have at least partial promoter activity (Table 3). The diversity is especially large at positions −11 to −9 (Fig. 2B; Table 1). Thus, promoter recognition might not require strict base-specific interactions at these positions. Instead, positions −11 to −9 may fulfill some RNA structural role that can accommodate a considerable amount of base substitutions, with a more gradual degradation of promoter recognition efficiencies.

Among alphaviruses, there is only a single difference, A-9 versus G-9, in the region from −13 to −9, and this conservation has been attributed to a combination of selection for promoter function and selection for nsP4 coding (2, 11). The present results confirm this, and they further delineate the relative contributions of each of the selective forces at different positions in the promoter. The STR promoter sequence, unlike the SIN wild-type promoter, does not overlap the nsP4 coding region and is not constrained by nsP4 coding requirements. Therefore, the sequences selected for are selected on the basis of promoter function alone. Positions −13 and −12 normally correspond to the second and third positions of an invariant tyrosine codon, requiring a U and a purine, respectively, on the minus strand. However, even in the absence of nsP4 coding constraints, the consensus sequences derived from the popu-

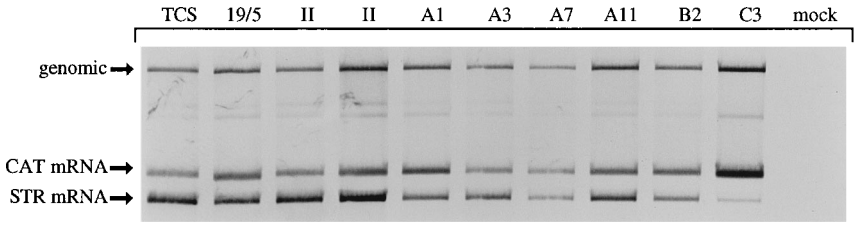

FIG. 4. Intracellular viral RNA synthesized by each virus. Cells were infected with each virus at an MOI of 5. Dactinomycin was added 1 h before labeling the cells with [^{3}H]uridine for 2 h. Intracellular RNA isolated at 5 h was denatured and separated through 1% agarose. The gel was treated for fluorography and exposed to film.

lation as a whole (Fig. 2B) and from the individual isolates (Table 2) show that the wild-type nucleotides are favored. Therefore, the sequence at positions −13 and −12 is primarily constrained by promoter activity. In contrast, non-wild-type bases are better tolerated at positions −11 to −9 (Fig. 2B; Table 2), which correspond to a glycine codon. This suggests that the conservation of the wild-type sequence 3′CCA on the minus strand is partly due to the amino acid coding on the plus strand. Most of the promoter sequences isolated (Table 1) would result in replacement of the glycine codon if they were placed in the normal context of the virus. For example, none of the 10 non-wild-type promoter sequences isolated from the P4 population encode the wild-type tyrosine-glycine nsP4 sequence. It is possible that at least some of these nsP4 changes are deleterious and would be eliminated during evolution.

Since the sequence of a virus population as a whole is a good representation of the promoters in the population, it provides a rapid assay to look for selection of sequences under a variety of conditions in this and other regions of the viral genome. This will facilitate the identification of the environmental factors important for selection and of the sequences that are optimal under each selection regime. For instance, the use of different host cells (5, 19) or passaging in vivo could result in different spectra of selected promoters.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI26763.

We thank Richard Pierce, Burton Wice, and William Parks for advice and encouragement throughout this project. We thank Charlie Rice, Sondra Schlesinger, and the SIN virologists at Washington University for helpful discussions and critical reading of the manuscript. We are grateful to Alan Templeton for help with the statistical analyses.

REFERENCES

1. **Carmichael, G. G., and G. K. McMaster.** 1980. The analysis of nucleic acids in gels using glyoxal and acridine orange. Methods Enzymol. **65:**380–391.
2. **Grakoui, A., R. Levis, R. Raju, H. V. Huang, and C. M. Rice.** 1989. A *cis*-acting mutation in the Sindbis virus junction region which affects subgenomic RNA synthesis. J. Virol. **63:**5216–5227.
3. **Hardy, W. R., and J. H. Strauss.** 1988. Processing the nonstructural polyproteins of Sindbis virus: study of the kinetics in vivo by using monospecific antibodies. J. Virol. **62:**998–1007.
4. **Hertz, J. M., and H. V. Huang.** 1992. Utilization of heterologous alphavirus junction sequences as promoters by Sindbis virus. J. Virol. **66:**857–864.
5. **Hertz, J. M., and H. V. Huang.** 1995. Host-dependent evolution of the Sindbis virus promoter for subgenomic mRNA synthesis. J. Virol. **69:**7775–7781.
6. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. **56:**335–341.
7. **Leitner, T., E. Halapi, G. Scarlatte, P. Rossi, J. Albert, E.-M. Fenyö, and M. Uhlen.** 1993. Analysis of heterogeneous viral populations by direct DNA sequencing. BioTechniques **15:**120–127.
8. **Levis, R., S. Schlesinger, and H. V. Huang.** 1990. Promoter for Sindbis virus RNA-dependent subgenomic RNA transcription. J. Virol. **64:**1726–1733.
9. **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.
10. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. **Ou, J.-H., C. M. Rice, L. Dalgarno, E. G. Strauss, and J. H. Strauss.** 1982. Sequence studies of several alphavirus genomic RNAs in the region containing the start of the subgenomic RNA. Proc. Natl. Acad. Sci. USA **79:**5235–5239.
12. **Raju, R., and H. V. Huang.** 1991. Analysis of Sindbis virus promoter recognition in vivo, using novel vectors with two subgenomic mRNA promoters. J. Virol. **65:**2501–2510.
13. **Raju, R., and H. V. Huang.** 1992. Unpublished data.
14. **Rice, C. M., R. Levis, J. H. Strauss, and H. V. Huang.** 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutants. J. Virol. **61:**3809–3819.
15. **Rice, C. M., and J. H. Strauss.** 1981. Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Natl. Acad. Sci. USA **78:**2062–2066.
16. **Roff, D. A., and P. Bentzen.** 1989. The statistical analysis of mitochondrial DNA polymorphisms: chi^2 and the problem of small samples. Mol. Biol. Evol. **6:**539–545.
17. **Smith, D. B., and S. C. Inglis.** 1987. The mutation rate and variability of eukaryotic viruses: an analytical review. J. Gen. Virol. **68:**2729–2740.
18. **Sokal, R. R., and F. J. Rohlf.** 1981. Biometry, p. 691–715. *In* J. Wilson (ed.), Biometry. W. H. Freeman and Company, San Francisco.
19. **Steinhauer, D. A., and J. J. Holland.** 1987. Rapid evolution of RNA viruses. Annu. Rev. Microbiol. **41:**409–433.
20. **Strauss, E. G., and J. H. Strauss.** 1986. Structure and replication of the alphavirus genome, p. 35–90. *In* S. Schlesinger and M. J. Schlesinger (ed.), The Togaviridae and Flaviviridae. Plenum Press, New York.
21. **Strauss, E. G., and J. H. Strauss.** 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. **58:**491–562.