Nucleic Acids Research, 1993, Vol. 21, No. 15 3529–3535

Prominent polypurine and polypyrimidine tracts in plant viroids and in RNA of the human hepatitis delta agent

Andrea D.Branch, Susan E.Lee, Olivia D.Neel and Hugh D.Robertson
Center for Studies of the Biological Correlates of Addiction, The Rockefeller University and The Cornell University Medical College, New York, NY 10021, USA

Received March 17, 1993; Revised and Accepted June 17, 1993

ABSTRACT

To seek patterns of nucleotide usage in the three types of circular subviral RNA pathogens, trimer frequencies and nearest-neighbor biases were studied in 12 plant viroid sequences; five sequences of circular plant viral satellite RNAs; and the sequence of RNA from the human hepatitis delta agent. The viroids and RNA of the delta agent contain tracts of polypurines and polypyrimidines which make up substantial portions of their genomes. Such tracts are not common in the virusoids or in the satellite RNA of tobacco ringspot virus. Viroids, the delta hepatitis agent, and the circular satellite RNAs of certain plant viruses have several features in common: all have circular genomic RNA and replicate through an RNA to RNA rolling circle replication cycle. However, virusoids and related satellite RNAs are directly or indirectly dependent on their helper viruses for replication, while the delta agent and viroids are not. The difference in the pattern of nucleotide usage between the plant viral satellite RNAs on the one hand, and viroids and delta RNA on the other, may relate to this difference in replication strategy.

INTRODUCTION

The oligonucleotide composition of genomic sequences varies markedly from organism to organism and often deviates strongly from the distribution of oligonucleotides found in random sequences of the same base composition. Analysis of oligonucleotide patterns has provided useful information about both entire genomes and subgenomic domains. For example, (i) the non-random representation of oligonucleotides, especially the under-representation of CTAG, led to a novel mutagenesis–recombination model for *E.coli* K-12 (1); (ii) the altered structure of both A-tract DNA (2,3) and Z-DNA (4) results from the presence of short sequence elements; (iii) strong correlations in the frequencies of oligonucleotides have shed light on the evolutionary relationships between cellular and viral sequences (5,6); and finally, (iv) nucleotide analysis has revealed the periodicity of several viroid sequences (7). Given the importance and variety of information that has resulted from studies of other systems, we investigated the patterns of oligonucleotides in the group of subviral pathogens which have circular RNA as their template for replication (8,9). These pathogens are all thought to replicate through an RNA to RNA rolling circle replication cycle (10) that generates precursor forms requiring RNA cleavage for maturation.

The circular subviral RNA pathogens can be divided into three groups: viroids, which are not associated with helper viruses (11); virusoids and related satellite RNAs of plants, which have helper viruses with RNA genomes (12–14); and the human hepatitis delta agent (15–18), whose helper virus, hepatitis B virus, replicates through reverse transcription of its pregenomic RNA (19). The delta agent is the only member of the circular subviral pathogens to have a host species that is not a higher plant and is also the only one known to specify a protein (17–18,20). In this report, we show that despite the evolutionary distance separating their host species, the human hepatitis delta agent and plant viroids both have extensive polypurine and polypyrimidine tracts which comprise major portions of their total length. The virusoids and related satellite RNAs associated with certain plant viruses do not have this feature.

METHODS

The pextran program, a component of the ARP software package developed at the Biomathematics Computation Laboratory of UCSF, was used for trimer analysis. Trimers were chosen for our study because of the extensive literature available concerning the trimer frequencies in other organisms. To analyze nearest-neighbor patterns, each nucleotide was classified as either matching the purine-pyrimidine class of its 5′ neighbor, or differing from its 5′ neighbor. In the sequence fragment Ca**A-GGAG**ug**A**, nucleotides in 'matching' tracts are in bold upper case letters and those in 'alternating' tracts are in lower case letters. While this fragment contains 6 consecutive purines, the longest (matching) tract is only 5 nucleotides long (because the first purine of the tract is preceded by a pyrimidine nucleotide). Any tract five or more nucleotides in length was designated a 'long tract.' Each nucleotide of a circular RNA was designated as a member of either a matching tract or an alternating tract. The 5′ terminal nucleotide of a linear RNA could not be assigned. Programs written by David Schreiber were used in our computer-assisted sequence analysis and are available upon request.

Table 1. Trimer frequencies of circular subviral RNA pathogens

I. PuPuPu&PyPyPy

	Viroids	Virusoids & Circ. Sat.	Delta Agent	X & Y RNAs	Delta Total	Delta Conserved
AAA	98	12	22	11	219	5
AAG	82	19	30	2	317	7
GAA	102	19	35	4	376	8
GAG	109	37	72	0	689	22
AGA	75	23	43	1	431	11
AGG	91	36	47	2	450	12
GGA	107	38	63	2	593	13
GGG	95	33	60	1	640	12
UUU	80	20	16	11	169	2
UUC	107	31	46	4	511	15
CUU	110	21	43	2	459	14
CUC	102	35	63	0	594	21
UCU	92	35	49	1	467	15
UCC	102	22	68	2	706	26
CCU	125	26	48	2	507	15
CCC	111	24	75	1	723	16
% of total	40%	25%	46%	37%	47%	45%

II. PuPyPu&PyPuPy

	Viroids	Virusoids & Circ. Sat.	Delta Agent	X & Y RNAs	Delta Total	Delta Conserved
ACA	26	18	7	0	87	1
ACG	33	31	11	0	122	0
GCA	42	15	12	0	115	2
GCG	74	27	22	0	225	7
AUA	13	17	5	0	47	0
AUG	13	17	17	0	178	7
GUA	22	30	10	5	86	1
GUG	57	35	15	0	159	4
CAC	46	34	15	0	142	2
CAU	9	23	13	0	152	6
UAC	28	35	6	5	59	1
UAU	18	11	6	0	66	1
CGC	81	25	19	0	195	7
CGU	43	30	10	0	114	3
UGC	49	14	16	0	152	6
UGU	42	34	9	0	93	2
% of total	15%	23%	11%	8%	12%	10%

III. PuPuPy&PyPyPu

	Viroids	Virusoids & Circ. Sat.	Delta Agent	X & Y RNAs	Delta Total	Delta Conserved
AAC	61	13	15	0	160	2
AGC	73	22	26	1	258	6
GAC	53	37	29	0	282	9
GGC	66	38	29	1	282	14
AAU	18	10	12	8	160	3
AGU	41	32	18	4	184	5
GAU	36	35	25	2	220	7
GGU	77	27	25	1	260	7
CCA	39	29	32	2	311	10
CUA	41	29	10	5	106	4
UCA	37	39	11	3	124	4
UUA	29	6	7	2	75	1
CCG	85	30	48	0	456	19
CUG	99	40	24	1	228	5
UCG	84	27	32	0	301	9
UUG	47	30	10	4	114	1
% of total	23%	26%	21%	27%	21%	22%

IV. PuPyPy&PyPuPu

	Viroids	Virusoids & Circ. Sat.	Delta Agent	X & Y RNAs	Delta Total	Delta Conserved
ACC	73	33	26	1	248	7
ACU	56	37	21	4	186	3
AUC	37	25	27	2	232	7
AUU	18	20	7	8	95	1
GCC	74	30	34	1	320	12
GCU	79	27	22	1	227	10
GUC	69	32	24	0	261	9
GUU	55	26	13	0	146	4
CAA	37	14	16	4	152	1
CAG	52	30	18	1	191	7
CGA	69	29	40	0	398	17
CGG	83	31	44	0	397	13
UAA	22	9	6	2	63	2
UAG	37	27	14	5	126	4
UGA	49	38	15	3	145	3
UGG	76	36	26	2	289	10
% of total	22%	26%	21%	27%	21%	22%

Trimer frequencies were determined for 12 viroid sequences (21–32); four virusoids (34–37) and sTobRV(38), the delta isolate sequenced by Wang *et al.* (15–16); X and Y RNAs of Konarska and Sharp (62); a total of 10 complete sequences of delta RNA (15–16,48–55); and 52 invariant sequences from delta RNA that were present in all 10 complete sequences of delta and had a minimum length of six nucleotides. Trimers were divided into four groups based on their purine-pyrimidine structure: Group I (PuPuPu&PyPyPy), Group II (PuPyPu&PyPuPy), Group III (PuPuPy&PyPyPu), Group IV (PuPyPy&PyPuPu). Finally, the trimers in each group were expressed as a percentage of the total.

RESULTS

Trimer frequencies

To begin an investigation of possible non-random patterns of oligonucleotides in the circular subviral RNA pathogens, we surveyed the distribution of trimers in 12 viroids (21–32), representing all three major viroid families (33); 5 circular satellite RNAs associated with plant viruses [four virusoids (13,34–37), and the satellite of tobacco ringspot virus (sTobRV), which is packaged as a linear RNA, but uses a circular template for replication (14,38–39)], and the delta agent (15,16). All these RNAs contain an almost equal percentage of purine and pyrimidine bases.

Both the viroids and the delta agent have an excess of trimers composed exclusively of either purines (Pu) or pyrimidines (Py). This group of trimers is abbreviated PuPuPu&PyPyPy. While this group of trimers has 16 members and thus makes up 25% of the total population of trimers, it constitutes 40% of the trimers present in viroid genomes and 46% of the trimers present in the

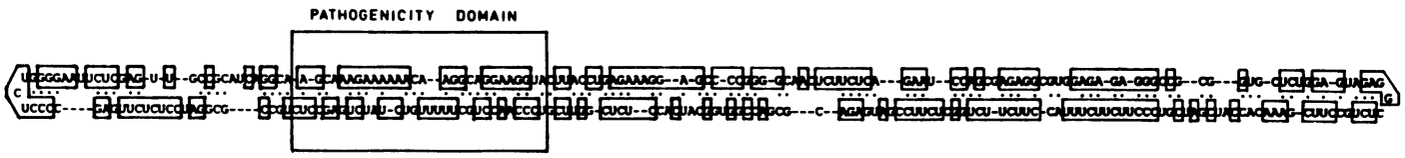

Figure 1. The rodlike secondary structure of HSV, with nucleotides in matching tracts enclosed in boxes.

delta agent (see Table 1). In contrast, the group of trimers with alternating purines and pyrimidines (which are abbreviated PuPyPu&PyPuPy and which comprise 25% of all possible trimers) constitutes only 15% of the trimers present in the viroid sequences and 11% of the delta agent. The trimer frequencies of the virusoids and related satellite RNAs differed from those of the viroids and the delta agent, with the PuPuPu&PyPyPy group of trimers constituting 25% of the total and the PuPyPu&PyPuPy group constituting 23% (Table 1). These data led to a more detailed nearest-neighbor analysis.

Polypurine and polypyrimidine tracts

To identify the pattern of nucleotides in the genomes of the circular subviral RNA pathogens, each nucleotide was classified as either matching or differing from the purine-pyrimidine class of its 5′ neighbor. Nucleotides making up matching tracts are enclosed in boxes in Figure 1, which depicts hop stunt viroid (HSV) in the viroid rodlike structure (40,41). In HSV, nucleotides in matching tracts make up 65% of the genome. Furthermore, 14 matching tracts containing 5 or more nucleotides (long matching tracts) occur in HSV, accounting for 34% of the genome. In contrast, HSV contains no long tracts of alternating purines and pyrimidines. Hence, such long alternating tracts make up 0% of the HSV genome (see Table 2).

Long matching tracts were present in all 12 viroid sequences analyzed. In the potato spindle tuber viroid (PSTV) family, they comprise from 34% of the genome, in the case of HSV, to 26%, in the case of PSTV. In the apple scar skin viroid (ASSV) family, the percentage of nucleotides in long matching tracts ranges from 27%, in pear blister canker viroid (PBCV) and grapevine yellow speckle viroid-2 (GYSV-2) to 17% in ASSV. In the two viroids capable of forming hammerhead self-cleavage structures (31,42–44), long matching tracts make up from 27%, in avocado sunblotch viroid (ASBV), to 17% in peach latent mottle viroid (PLMV) (see Table 2). In comparison to the prominence of long matching tracts, long alternating tracts comprised a small percentage of the viroid genomes we analyzed. They accounted for a maximum of 7%, in coconut cadang cadang viroid (CCCV), an RNA in which long matching tracts make up 32% of the sequence (Table 2).

One of the polypurine tracts in viroid RNA has been noted many times previously, having been identified by Gross *et al.* in the first complete viroid sequence (25). It is a hallmark feature of the pathogenesis domain as defined by Keese and Symons (45) and constitutes much of the upper strand of this domain in both PSTV-like and ASSV-like viroids (33; see Figure 1). This polypurine tract is undoubtedly an important element, in and of itself, and as our study shows, it reflects a larger trend. In every case, the nucleotides in this tract are a minority of the total present in long matching tracts.

Unlike the viroids, the virusoids and sTobRV have no bias favoring long matching tracts over long alternating tracts. In this group of infectious RNAs, long matching tracts account for less than 10% of any genome and long alternating tracts account for a similar and small percentage (Table 2). Recently, the complete sequence of a small circular RNA from carnation plants was reported by Hernández *et al.* (46). This RNA has a distribution of trimers similar to that of the virusoids and sTobRV. It has no long alternating tracts and only one long matching tract, which comprises just 7% of the genome. These features suggest that it may in someway be functionally related to the virusoids and sTobRV. We have not included a detailed analysis of this RNA, carnation stunt associated viroid, because 'both its viroid nature and causal relationship with the stunting syndrome in carnation remain to be fully demonstrated' (46).

Clusters of purines and clusters of pyrimidines dominate the sequence of delta RNA. In this way, the delta agent is more similar to viroids than to the circular satellite RNAs of plants. In the 1679 base-long sequence of delta RNA determined by Wang *et al.* (15–16), there are 77 segments each containing five or more nucleotides which match their 5′ neighbor (see Table 2). Such long matching tracts are boxed in Figure 2, which presents the rodlike secondary structure of delta RNA worked out by Wang *et al.* (15). Long matching tracts contain a total of 594 nucleotides; account for 35% of the total length of delta RNA; and are prominent throughout the molecule, suggesting that they represent a ubiquitous structural element which may be important for replication and propagation. In contrast, there are only two long alternating tracts in delta RNA. They are underlined in Figure 2. These two long alternating tracts contain only 12 nucleotides in total, and comprise just 0.7% of the genome (in Table 2, this number is rounded off to 1%). In the viroid-like domain, residues 613–980 (see Figure 2 and ref. 47), 26% of the nucleotides are in long matching tracts; while only 2% of the nucleotides are in long alternating tracts. In the protein-coding domain, which makes up the remainder of the molecule, 38% of the residues are in long matching tracts and only 0.4% are in long alternating tracts.

To see whether the highly conserved sequences in delta RNA show the same nearest-neighbor effects as the molecule as a whole, the trimer frequencies of ten complete delta sequences (15–16,48–55) were determined and are presented in Table 1 ('Delta Total'). These values were compared to the trimer frequencies of highly conserved sequences, which are also presented in Table 1 ('Delta Conserved'). In this analysis, highly conserved sequences were defined as sequences at least six nucleotides in length present in every isolate. There are 52 such sequences, constituting 35% of the total genome and comprising 64% of the viroid-like domain, (emphasizing the high degree of sequence conservation of this domain). The group of trimers composed exclusively of purines or exclusively of pyrimidines (denoted PuPuPu&PyPyPy in Table 1) make up 47% of the total delta genome and 45% of the invariant sequences; while the class of trimers composed of alternating purine and pyrimidine

Table 2. Lengths of matching and alternating tracts in the circular subviral RNA pathogens

TALLY OF MATCHING TRACTS OF A GIVEN LENGTH

	Short Tracts				Long Tracts														Long
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
HSV	21	13	10	4	4	3	1	2	3			1							34%
CCCV	21	13	4	3	2	5	2	2	1										32%
CSV	30	9	16	10	4	3	3		1	1		1		1					29%
CEV	29	15	14	9	3	3	4	1		1		1					1		29%
PSTV	29	14	12	11	4	2	1	2	1				1				1		26%
PBCV	27	16	8	7	6	3		1	1	1	1								27%
GYSV-2	25	19	13	9	3	2	3	3			1					1			27%
GYSV-1	30	21	12	9	3	4	1		1	1		1							21%
AGV	40	16	13	8	5	3	2			1			1						21%
ASSV	36	18	8	6	2	1	1		1		1			1					17%
ASBV	19	12	11	4	3	4	1	1					1						27%
PLMV	27	22	17	5	2	3				1									17%
Delta	125	76	38	37	19	13	11	16	5	4	1	2	1		2		1	2	35%
vSNMoV	51	27	8	7	2	3	1												9%
vLTSVA	39	26	15	4	1	3	1												9%
vLTSVc	43	21	19	1	2	2													7%
vSCMoV	36	21	12	6	3		1												7%
sTobRV	48	18	9	10	2	2													6%

TALLY OF ALTERNATING TRACTS OF A GIVEN LENGTH

	Short Tracts				Long Tracts														Long
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
HSV	32	21	6	3															0%
CCCV	31	11	7	1	2			1											7%
CSV	51	18	8		1														1%
CEV	47	25	7	2															0%
PSTV	42	25	8	3															0%
PBCV	40	19	10	1		1													1%
GYSV-2	50	16	9	3	1														1%
GYSV-1	47	19	9	6	2														3%
AGV	51	23	12	2	1														1%
ASSV	27	31	8	7	1	1													3%
ASBV	31	19	5	1															0%
PLMV	44	21	11	3	1														1%
Delta	215	96	33	7	1		1												1%
vSNMoV	53	26	10	3	4	2	1												10%
vLTSVA	53	27	5	1	2	1													5%
vLTSVc	50	23	7	5	1	2													5%
vSCMoV	34	22	9	8	3	2		1											11%
sTobRV	43	24	5	9	6		2												12%

Each nucleotide was categorized as belonging to a matching or an alternating sequence tract, based on identity with, or difference from, the purine-pyrimidine class of its 5′ neighbor. The total number of tracts for each length were counted. The percentage of an RNA sequence in long tracts, i.e., tracts five or more nucleotides in length, was calculated. The following RNAs were analyzed: hop stunt viroid (HSV) (21); coconut cadang cadang viroid (CCCV) (22); chrysanthemum stunt viroid (CSV) (23); citrus exocortis viroid (CEV) (24); potato spindle tuber viroid (PSTV) (25); pear blister canker viroid (PBCV) (26); grapevine yellow speckle viroid-2 (GYSV-2) (27); grapevine yellow speckle viroid-1 (GYSV-1) (28); Australian grapevine viroid (AGV) (29); apple scar skin viroid (ASSV) (30); avocado sunblotch viroid (ASBV) (31); peach latent mottle viroid (PLMV) (32); the delta hepatitis agent (15–16); the virusoid of solanum nodiflorum mottle virus (vSNMoV) (34); the virusoid of lucerne transient streak virus from Australia (vLTSVA) (35); the virusoid of lucerne transient streak virus from Canada (vLTSVC) (36); the virusoid of subterranean clover mottle virus (vSCMoV) (37); the satellite of tobacco ringspot virus (sTobRV) (38).

nucleotides make up 12% and 10% of the total sequences and the invariant sequences, respectively (Table 1). Thus, the highly conserved sequences within delta RNA have a pattern of oligonucleotides similar to that of the genome as a whole.

To complete our study, two types of random sequences were analyzed. In three random sequences having the size and base composition of delta RNA, 8% of the nucleotides were in long matching tracts and 10% were in long alternating tracts. This pattern contrasts sharply with that of delta RNA itself, in which long matching tracts comprise 35% of the genome and long alternating tracts comprise 0.7%. In three random sequences with the base composition and the approximate size of the circular plant viral satellite RNAs, the percentages of nucleotides in long matching tracts and in long alternating tracts were 11% and 8%, respectively. These percentages correspond closely with those observed in their natural counterparts, in which nucleotides in long matching tracts comprise an average of 8% and those in long alternating tracts comprise an average of 9% of the total.

DISCUSSION

Summary

Our analysis reveals an underlying similarity in the primary structure of the delta agent (15–16) and viroids (33): the delta agent and all three categories of viroids (PSTV-like viroids, ASSV-like viroids, and the two viroids containing hammerhead

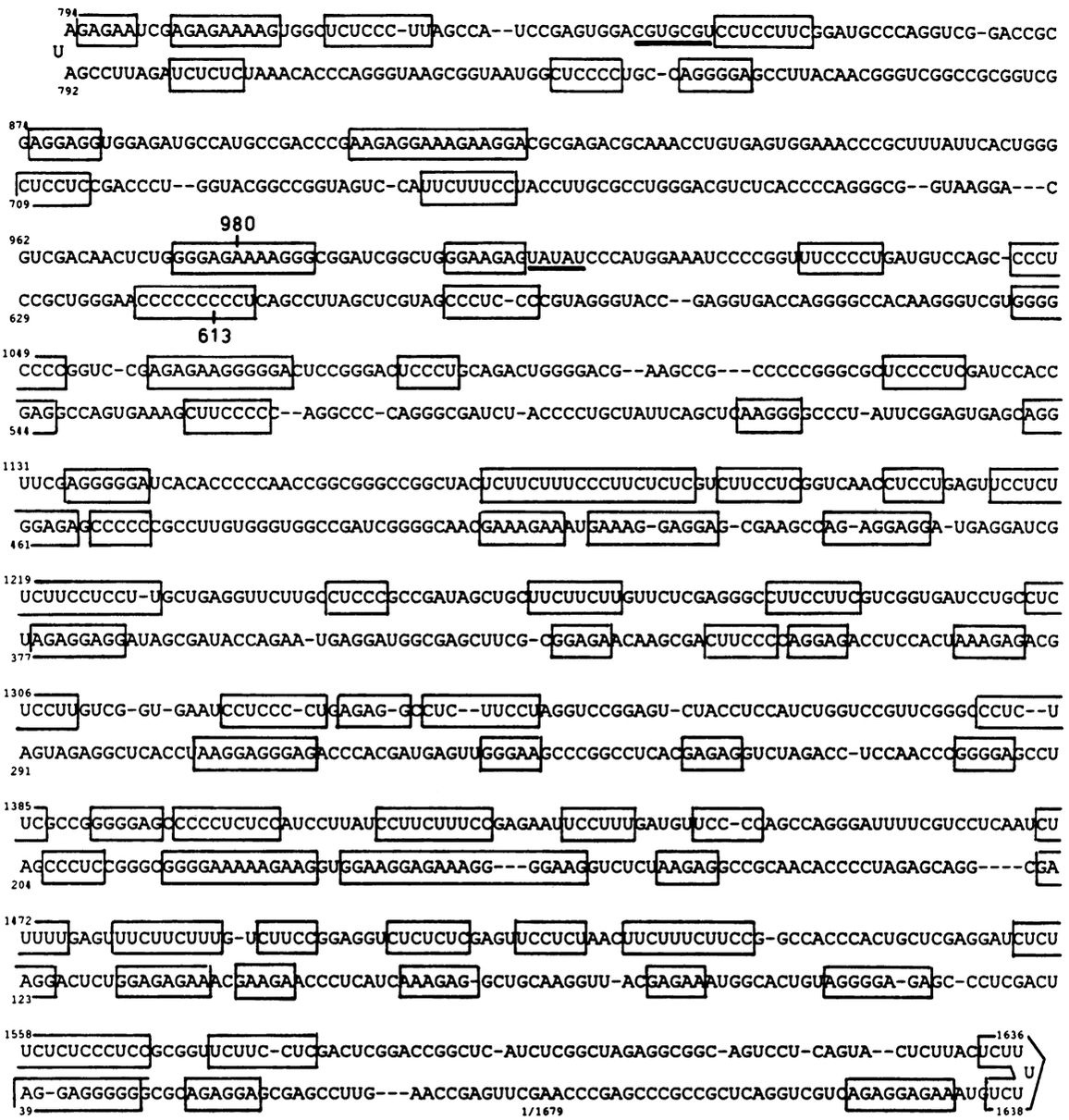

Figure 2. The rodlike secondary structure map of delta RNA, with nucleotides in long matching tracts enclosed in boxes.

motifs) have a dearth of sequences more than a few nucleotides in length in which purine and pyrimidine nucleotides alternate, while they have many polypurine and polypyrimidine tracts. This pattern of nucleotides is not evident in the virusoids and sTobRV (12–14), which are members of the third major group of the circular subviral RNA pathogens (8). While we do not know the significance of the polypurine and polypyrimidine tracts in viroids and delta RNA, we wish to consider three processes which may be related to this aspect of their primary structures.

Possible significance of the polypurine and polypyrimidine tracts

First, viroids and the delta agent can be copied in virus-free cells (11,56), unlike RNAs such as the virusoids and sTobRV, which are somehow dependent upon their helper viruses for replication and may use viral components for some steps in this process (12–14,57). A growing body of evidence suggests that the viroids, PSTV (58–59) and citrus exocortis viroid (CEV) (60), and the delta agent (61) are all copied by RNA polymerase II of their hosts, although it is not known what features of viroids and delta RNA allow them to be replicated by these enzymes. Perhaps significantly, the X and Y RNAs described by Konarska and Sharp (62), which are also copied by a DNA-dependent RNA polymerase (that of bacteriophage T7) have a pattern of trimer frequencies similar to that of viroids and the delta agent. In the X and Y RNAs, 37% of the trimers are of the PuPuPu&PyPyPy category and only 8% are of the PuPyPu&PyPuPy category (see Table 1), suggesting that a segregation of purine and pyrimidine nucleotides may be a recurring feature of RNAs that are replicated by enzymes normally involved in the transcription of DNA.

Second, PSTV and delta RNA not only appear to be copied by RNA polymerase II, they also accumulate to high concentrations in the same cellular compartment, the nucleus (63–65). In contrast, there is evidence that the virusoid of velvet

tobacco mottle virus replicates in the cytoplasm (66). Survival in the nucleus may impose constraints which are reflected in the sequence differences distinguishing viroids and delta, on the one hand, and circular satellite RNAs associated with plant viruses, on the other.

Third, the physical properties of the viroids and virusoids differ in ways which have been difficult to explain and which may relate to the primary sequence biases we have identified. For example, CEV and the virusoid of solanum nodiflorum mottle virus (vSNMoV) are nearly identical in length (371 and 377 nucleotides, respectively); percentage of the nucleotides which are base paired (68% and 66%, respectively); and the percentage of the base pairs which are GC basepairs (59% and 57%, respectively). However, the T_m values of these two circular RNAs differ by 13°C (67). Through studies of model RNA helices composed of sequences derived from viroid and virusoid sequences, it may be possible to determine whether the much greater stability of viroid RNA is related to its tendency to segregate purine and pyrimidine nucleotides.

Finally, the polypurine and polypyrimidine tracts may advance our understanding of the viroid family tree. The underlying similarity among viroid sequences, and the limited complexity of these sequences, suggest that convergent evolution may play a larger role in the generation of certain new viroids than has been previously recognized. For example, while columnea latent viroid (CLV) is considered to be a chimeric product of intracellular recombination between an HSV-type and PSTV-type viroid, with residues 96–110 and 261–281 coming from the HSV-type parent (68), in fact, PSTV, CLV and HSV are all quite similar to each other in these regions and could plausibly be derived from a single common ancestral sequence through mutation and natural selection, without an RNA recombination event:

```
GGAUCCCCGGGGAAACC (PSTV)
                  UAC-U-ACCCGGUGGAAACAACU (PSTV)
GGAGCCCCGGGGCAACU (HSV&CLV)
                  GACGCGA-CCGGUGGCAUC-ACC (HSV&CLV)
GGA-CCCCGGGG-AAC- (invariant)
                  -AC---A-CCGGUGG-A-C-AC- (invariant)
```

The short identical sequences in HSV and CLV may have arisen independently in HSV and CLV through a two-step process involving an error-prone polymerase reaction and a selection process which rejects most sequences as nonviable. The power of these two forces acting in concert was recently demonstrated by Lakshman and Tavantzis, who introduced a two-base deletion into a PSTV cDNA sequence, only to recover wild-type PSTV which had been generated *de novo* in host plants (69). This result underscores the earlier work by Owens and colleagues showing that wild-type PSTV was synthesized in plants inoculated with a PSTV cDNA clone containing a single nucleotide substitution (70).

The regenerative capacity of the viroid's mutation-selection-amplification process illustrates the potential speed of convergent viroid evolution. Moreover, Glenn *et al.* found that flanking vector sequences were neatly deleted from transcripts of delta RNA during the replication-selection cycle taking place in transfected cells (71), revealing the regenerative capacity of this system, as well. As more sequences of circular subviral RNA pathogens become available for analysis, the products of RNA recombination and the products of convergent evolution may become easier to identify with certainty.

CONCLUSIONS

Compared to the majority of micro-organisms, viroids, virusoids, and the delta agent all have small genomes. In some cases, a variety of diverse functions have been mapped onto these genomes. For example, one segment of delta's genomic RNA can form a ribozyme (72–75); a second is the template for a related ribozyme (76); a third is the substrate for an RNA editing event (77–78); a fourth folds into a UV-sensitive tertiary structure (47); a fifth is the template for a functional open reading frame (20). The entire molecule is the template for RNA to RNA rolling circle replication (10,79) and the mature RNA contains binding sites for the delta antigen (80–81). Finally, recent unpublished experiments carried out in collaboration with M.B. Mathews and his colleagues (Cold Spring Harbor) indicate that delta RNA transcripts have novel structures recognized by an antiviral cellular kinase (82). Considering its need to accommodate many diverse functions within a small genome, we were surprised to find that delta RNA has long tracts of polypurines and polypyrimidines. This pattern is evident throughout the entire molecule and thus occurs in both the viroid-like domain and the protein-coding domain (see Figure 2). In the future, we hope to learn the biological significance of this pattern of nucleotides and thus account for its presence in both the delta agent and in plant viroid RNAs.

ACKNOWLEDGMENTS

We thank Dave Schreiber and Jana Balak for their excellent assistance. This work was supported in part by a NIDA Research Center grant, P50DA-05130, which is directed by Dr Mary Jeanne Kreek of the Rockefeller University, and by grant U01AI-31867 from NIH.

REFERENCES

1. Merkl,R., Kröger,M., Rice,P. and Fritz,H.-J. (1992) *Nucleic Acids Res.*, **20**, 1657–1662.
2. Hagerman,P.J. (1990) *Annu. Rev. Biochem.*, **59**, 755–781.
3. Gartenberg,M.R. and Crothers,D.M. (1991) *J. Mol. Biol.*, **219**, 217–230.
4. Rich,A., Nordheim,A. and Wang,A.H.-J. (1984) *Ann. Rev. Biochem.*, **53**, 791–846.
5. Subak-Sharpe,H., Bürk,R.R., Crawford,L.V., Morrison,J.M., Hay,J. and Keir,H.M. (1966) *Cold Spring Harbor Symposia on Quantitative Biology*, **31**, 737–748.
6. Barrai,I., Scapoli,C., Barale,R. and Volinia,S. (1990) *Nucleic Acids Res.*, **18**, 3021–3025.
7. Juhász, A., Hegyi, H. and Solymosy, F. (1988) *Biochimicia et Biophysica Acta*, 950, 455–458.
8. Branch,A.D., Levine,B.J. and Robertson,H.D. (1990) *Semin. Virology*, **1**, 143–152.
9. Elena,S.F., Dopazo,J., Flores,R., Diener,T.O. and Moya,A. (1991) *Proc. Nat. Acad. Sci., U.S.A.*, **88**, 5631–5634.
10. Branch,A.D., Robertson,H.D. (1984) *Science*, **223**, 450–455.
11. Diener,T.O. (1985) The Viroids. Plenum Press, New York.
12. Francki,R.I.B. (1985) *Ann. Rev. Microbiol.*, **39**, 151–174.
13. Keese,P. and Symons,R.H. (1987) In Semancik,J.S. (ed.), Viroids and Viroid-Like Pathogens. CRC Press, Boca Raton, Florida, pp. 1–47.
14. Bruening,G. (1990) *Semin. Virology*, **1**, 127–134.
15. Wang,K.-S., Choo,Q.-L., Weiner,A.J., Ou,J.-H., Najarian,R.C., Thayer,R.M., Mullenbach,G.T., Denniston,K.J., Gerin,J.L and Houghton,M. (1986) *Nature*, **323**, 508–514.
16. Wang,K.-S., Choo,Q.-L., Weiner,A.J., Ou,J.-H., Najarian,R.C., Thayer,R.M., Mullenbach,G.T., Denniston,K.J., Gerin,J.L and Houghton,M. (1986) *Nature*, **328**, 456.
17. Rizzetto,M. (1990) *J. Hepatology*, **11**, S145–S148.

18. Gerin,J.L., Purcell,R.H. and Rizzetto,M. (ed.) (1991) The Hepatitis Delta Virus. Wiley-Liss,Inc., New York, Vol. 364.
19. Mason,W.S., Lien,J.-M., Petcu,J., Coates,L., London,W.T., O'Connell,A.P., Aldrich,C. and Custer,R.P. (1987) In Robinson,W.S., Koike,K. and Will,H. (ed.) Hepadna Viruses. Alan R. Liss,Inc., New York, pp.3–16.
20. Weiner,A.J., Choo,Q.-L., Wang,K.-S., Govindarjan,S., Redeker, A.G., Gerin,J.L. and Houghton, M. (1988) *J Virol.*, **62**, 594–599.
21. Ohno,T., Takamatsu,N., Meshi,T. and Okada,Y. (1983) *Nucleic Acids Res.*, **11**, 6185–6197.
22. Haseloff,J., Mohamed,N.A. and Symons,R.H. (1982) *Nature*, **299**, 316–321.
23. Haseloff,J. and Symons,R.H. (1981) *Nucleic Acids Res.*, **9**, 2741–2752.
24. Gross,H.J., Krupp,G., Domdey,H., Raba,M., Jank,P., Lossow,C., Alberty,H., Ramm,K. and Sänger,H.L. (1982) *Eur. J. Biochem.*, **121**, 249–257.
25. Gross,H.J., Domdey,H., Lossow,C., Jank,P., Raba,M., Alberty,H. and Sänger,H.L. (1978) *Nature*, **273**, 203–208.
26. Hernández,C., Elena,S.F., Moya,A. and Fores,R. (1992) *J. Gen. Virology*, **73**, 2503–2507.
27. Koltunow,A.M. and Rezaian,M.A. (1989) *Virology*, **170**, 575–578.
28. Koltunow,A.M. and Rezaian,M.A. (1988) *Nucleic Acids Res.*, **16**, 849-864.
29. Rezaian,M.A. (1990) *Nucleic Acids Res.*, **18**, 1813–1818.
30. Puchta,H., Luckinger,R., Yang,X., Hadidi,A. and Sänger,H.L. (1990) *Plant Molecular Biology*, **14**, 1065–1067.
31. Symons,R.H. (1981) *Nucleic Acids Res.*, **23**, 6527–6537.
32. Hernández,C. and Flores,R. (1992) *Proc. Nat. Acad. Sci., U.S.A.*, **89**, 3711–3715.
33. Koltunow,A.M. and Rezaian,M.A. (1989) *Intervirology*, **30**, 194–201.
34. Haseloff,J. and Symons,R.H. (1982) *Nucleic Acids Res.*, **10**, 3681–3691.
35. Keese,P., Bruening,G. and Symons,R.H. (1983) *FEBS Lett.*, **159**, 185–190.
36. Abouhaidar,M.G. and Paliwal,Y.C. (1988) *J. Gen. Virology*, **69**, 2369–2373.
37. Davies,C., Haseloff,J. and Symons,R.H. (1990) *Virology*, **177**, 216-224.
38. Buzayan,J.M., Gerlach,W.L., Bruening,G., Keese,P. and Gould,A.R. (1986) *Virology*, **151**, 186–199.
39. Linthorst,H.J.M. and Kaper,J.M. (1984) *Virology*, **137**, 206–210.
40. Sogo,J.M., Koller,T.H. and Diener,T.O. (1973) *Virology*, **55**, 70–80.
41. Riesner,D., Henco,K., Rokohl,U., Klotz,G., Kleinschmidt,A.K.,Domdey,H., Jank,P., Gross,H.J. and Sänger,H.L. (1979) *J. Mol. Biol.*, **133**, 85–115.
42. Prody,G.A., Bakos,J.T., Buzayan,J.M., Schneider,I.R. and Bruening,G. (1986) *Science*, **231**, 1577–1580.
43. Hutchins,C.J., Rathjen,P.D., Forster,A.C. and Symons,R.H. (1986) *Nucleic Acids Res.*, **14**, 3627–3640.
44. Symons,R.H. (1992) *Annu. Rev. Biochem.*, **61**, 641–671.
45. Keese,P. and Symons,R.H. (1985) *Proc. Natl. Acad. Sci., U.S.A.*, **82**, 4582–4586.
46. Hernández,C., Daròs,J.A., Elena,S.F., Moya,A. and Flores,R. (1992) *Nucleic Acids Res.*, **20**, 6323–6329.
47. Branch,A.D., Benenfeld,B.J., Baroudy,B.M., Wells,F.V., Gerin,J.L. and Robertson,H.D. (1989) *Science*, **243**, 649–652.
48. Saldanha,J.A., Thomas,H.J. and Monjardino,J.P. (1990) *J. Gen. Virology*, **71**, 1603–1606.
49. Kuo,M.Y.P., Goldberg,J., Coates,L., Mason,W., Gerin,J. and Taylor,J. (1988) *J. Virology*, **62**, 1855–1861.
50. Chao,Y.-C., Lee,C.-M., Tang,H.-S., Govindarajan,S. and Lai,M.M.C. (1991) *Hepatology*, **13**, 345–352.
51. Chao,Y.-C., Chang,M.-F., Gust,I. and Lai,M.M.C. (1990) *Virology*, **178**, 384–392.
52. Makino,S., Chang,M.-F., Shieh,C.-K., Kamahora,T., Vannier,D.M., Govindarajan,S. and Lai,M.M.C. (1987) *Nature*, **329**, 343–346.
53. Imazeki,F., Omata,M. and Ohto,M. (1991) *Nucleic Acids Res.*, **19**, 5439.
54. Kos,T., Molijn,A., van Doorn,L.-J., van Belkum,A. and Schellekens,H. (1991) *J. Med. Virology*, **34**, 268–279.
55. Lee,C.-M., Bih,F.-Y., Chao,Y.-C., Govindarajan,S. and Lai,M.M. (1992) *Virology*, **188**, 265–273.
56. Taylor,J., Mason,W., Summers,J., Goldberg,J., Aldrich,C., Coates,L., Gerin,J. and Gowans,E. (1987) *J. Virology*, **61**, 2891–2895.
57. Schneider,I.R. (1971) *Virology*, **45**, 108–122.
58. Rackwitz,H.R., Rohde,W., Sänger,H.L. (1981) *Nature*, **291**, 297–301.
59. Schindler,I.-M. and Mühlbach,H.-P. (1992) *Plant Science*, **84**, 221–229.
60. Flores,R. and Semancik,J.S. (1982) *Proc. Natl. Acad. Sci., U.S.A.*, **79**, 6285–6288.
61. MacNaughton,T.B., Gowans,E.J., McNamara,S.P. and Burrell,C.J. (1991) *Virology*, **184**, 387–390.
62. Konarska,M.M. and Sharp,P.A. (1990) *Cell*, **63**, 609–618.
63. Takahashi,T. and Diener,T.O. (1975) *Virology*, **64**, 106–114.
64. Harders,J., Lukács,N., Robert-Nicoud,M., Jovin,T.M. and Riesner,D. (1989) *EMBO J.*, **8**, 3941–3949.
65. Gowans,E.J., Baroudy,B.M., Negro,F., Ponzetto,A., Purcell,R.H. and Gerin,J.L. (1988) *Virology*, **167**, 274–278.
66. Randles,J.W., Davies,C., Hatta,T., Gould,A.R. and Francki,R.I.B. (1981) *Virology*, **108**, 111–122.
67. Riesner,D., Colpan,M., Goodman,T.C., Nagel,L., Schumacher,J. and Steger,G. (1983) *J. Biomol. Struc. Dynam.*, **1**, 669–688.
68. Hammond,R., Smith,D.R. and Diener,T.O. (1989) *Nucleic Acids Res.*, **17**, 10083–10094.
69. Lakshman,D.K. and Tavantzis,S.M. (1992) *Virology*, **187**, 565–572.
70. Owens,R.A., Hammond,R.W., Gardner,R.C., Kiefer,M.C., Thompson,S.M. and Cress,D.E. (1986) *Plant Mol. Biol.*, 6, 179–192.
71. Glenn,J.S., Taylor,J.M. and White,J.M. (1990) *J. Virol.*, **64**, 3104–3107.
72. Kuo,M.Y.P., Sharmeen,L., Dinter-Gottlieb,G. and Taylor,J. (1988) *J. Virol.*, **62**, 4439–4444.
73. Wu,H.-N., Lin,Y-L., Lin,F-P, Makino,S., Chang,M.-F. and Lai,M.M.C. (1989) *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 1831–1835.
74. Perrotta,A.T. and Been,M.D. (1991) *Nature*, **350**, 434–436.
75. Branch,A.D. and Robertson,H.D. (1991) *Proc. Nat. Acad. Sci., U.S.A.*, **88**, 10163–10167.
76. Sharmeen,L., Kuo,M.Y.P., Dinter-Gottlieb,G. and Taylor,J. (1988) *J. Virol.*, **62**, 2674–2679.
77. Luo,G., Chao,M., Hsieh,S.-Y., Sureau,C., Nishikura,K. and Taylor, J. (1990) *J. Virol.*, **64**, 1021–1027.
78. Casey,J.L., Bergmann,K.F., Brown,T.L. and Gerin,J.L. (1992) *Proc. Nat. Acad. Sci., U.S.A.*, **89**, 7149–7153.
79. Chen,P.J., Kalpana,G., Goldberg,J., Mason,W., Werner,B., Gerin,J. and Taylor,J. (1986) *Proc. Nat. Acad. Sci., U.S.A.*, **83**, 8774–8778.
80. Lin,H.-J., Chang,M.-F., Baker,S.C., Govindarajan,S. and Lai,M.M.C. (1990) *J. Virol.*, **64**, 4051–4058.
81. Chao,M., Hsieh,S.-Y. and Taylor,J. (1991) *J. Virol.*, **65**, 4057–4062.
82. Kostura,M. and Mathews,M.B. (1989) *Mol. Cell. Biol.*, **9**, 1576–1586.