The Repeat Sequence of a Hermit Crab Satellite Deoxyribonucleic Acid is $(\text{-T-A-G-G-})_n \cdot (\text{-A-T-C-C-})_n$†

Dorothy M. Skinner,* Wanda G. Beattie, Frederick R. Blattner,‡
Barbara P. Stark, and James E. Dahlberg

ABSTRACT: We have separated the strands of hermit crab satellite I DNA, rich in deoxyguanylate and deoxycytidylate, and determined the sequence of each by analysis of their individual RNA transcripts. The tetramers -A-U-C-C- and -U-A-G-G- transcribed from the individual strands made up 86–92% of the RNA products. A minor sequence, of 4:1 (-C-U-G-):(-C-A-G-), was present at less than 5%. We believe this to be a different satellite DNA.

A naturally occurring satellite that makes up less than 0.2% of the DNA of the hermit crab, *Pagurus pollicaris*, has unusual physical and chemical properties which resemble those of simple alternating $(dG\text{-}dT)_n \cdot (dC\text{-}dA)_n$. It is one of a pair of "isopycnic twin" satellites, so-called because the two satellites form a single band at one density in neutral CsCl. Satellite I has been separated in its native form from its partner, satellite II, by centrifugation in Cs_2SO_4 gradients containing Hg^{2+} and Ag^+ (Skinner and Beattie, 1973). The strands of satellite I have been separated and characterized. Their base compositions are similar to those of the individual strands of the guinea pig α satellite (Corneo *et al.*, 1968), in that one strand has virtually no cytosine residues and the other strand no guanine residues (Skinner and Beattie, 1974). We describe here our studies of the sequence of this hermit crab satellite.

Sequence analysis of DNAs is greatly facilitated if they are relatively simple and small, and if their strands are separable (Barrell, 1971). Because of their simplicity and interstrand base compositional bias several satellite DNAs have been partially or completely sequenced. The sequences were determined from direct analyses of pyrimidine tracts of depurinated DNA (Southern, 1970), or by transcription into RNA (Fry *et al.*, 1973; Gall and Atherton, 1974), or *in vitro* DNA synthesis (Fry *et al.*, 1973). Partial sequences of other, more complicated DNAs have also been determined by direct analysis (Ziff *et al.*, 1973; Robertson *et al.*, 1973) and by DNA synthesis (Loewen and Khorana, 1973; Sanger *et al.*, 1973). In the present work, we have used transcription into RNA and sequence analysis of the transcript by chromatographic methods (Blattner and Dahlberg, 1972; Barrell, 1971; Brownlee and Sanger, 1967).

Materials and Methods

Template DNAs. DNA was isolated from epidermis, hemocytes, and testes of hermit crabs (*P. pollicaris*). Satellite DNAs (density 1.724 g/cm^3) were isolated by centrifugation of total DNA in neutral CsCl gradients. The isopycnic twin satellites I and II were separated from each other by centrifugation in $Hg^{2+} + Ag^+ - Cs_2SO_4$ gradients (Skinner and Beattie, 1973). Satellite I was checked for purity by thermal dissociation (Skinner and Beattie, 1974) and separated into G-rich (heavy, H)[1] and C-rich (light, L) strands in alkaline CsCl (Vinograd *et al.*, 1963).

Synthesis of Complementary RNA. The reaction mixture contained: 2.4 μg of DNA (single stranded); 1 nmol each of the nucleoside triphosphates (NTPs), one (or, where noted, all four) of which was labeled in the α position with ^{32}P (8.5–11.1 Ci/mmol); 200 nmol of $MgCl_2$; 1.2 μmol of Tris (pH 7.9), and 180 μg of *Escherichia coli* RNA polymerase in a total volume of 0.03 ml. RNA polymerase was isolated by the Chamberlin–Berg (1962) procedure. These concentrations were selected to facilitate the transcription of the single-stranded templates. The reaction was stopped by adding ethylenediaminetetraacetic acid (EDTA) to 400 mM. Incorporation was monitored by total counts and acid-precipitated counts collected on Whatman No. 3 filters (Bollum, 1959).

The RNA transcript was separated from unincorporated nucleotides by passage through a 42 × 0.9 cm column of Sephadex G-50 in the presence of 200 mM ammonium acetate (pH 7.9). Appropriate fractions were pooled, and 100 μg of carrier RNA was added. The pH was lowered to 6.5 with acetic acid and 5 mM $CaCl_2$ and 50 mM $MgCl_2$ (final concentrations) were added. "RNase-free DNase" from Worthington, treated with iodoacetate (Zimmerman and Sandeen, 1966), was added and incubated for 10 min at room temperature. The solution was then made 60 mM in EDTA, 0.125% in sodium dodecyl sulfate, and 0.2% in diethyl pyrocarbonate (Baycovin; Solymosy *et al.*, 1968) and shaken vigorously. After 10 min at 37° it was made 500 mM in KCl and evacuated to remove CO_2. The precipitated sodium dodecyl sulfate and proteins were removed by centrifugation in the cold. The RNA was then precipitated from the supernatant by addition of ethanol to 75%.

Ribonucleoside triphosphates were purchased from New England Nuclear Corp. and had specific activities of 8.5–11.1 Ci/mmol.

The purified RNA preparations were split in half for T_1

†From the Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830 (D.M.S. and W.G.B.), and from the McArdle Laboratory for Cancer Research (F.R.B.) and the Department of Physiological Chemistry (B.P.S. and J.E.D.), University of Wisconsin, Madison, Wisconsin 53706. *Received April 8, 1974.* This research was sponsored by the U. S. Atomic Energy Commission under contract with the Union Carbide Corporation and by National Science Foundation Grant 32152X to J.E.D. J.E.D. holds National Institutes of Health Research Career Development Award K-04-GM32770.

‡Present address: Department of Genetics, University of Wisconsin, Madison, Wis. 53706.

[1]Abbreviations used are: H strand, heavy (guanine-rich) strand; L strand, light (cytosine-rich) strand; NTP, nucleoside triphosphate; EDTA, ethylenediaminetetraaceticacid.

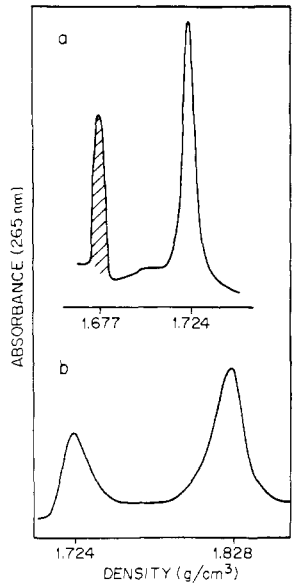

FIGURE 1: Tracing of Model E analytical ultracentrifuge pattern of satellite I in CsCl gradients: (a) 2 μg of DNA in neutral CsCl gradient. (the marker DNA (shaded peak) was poly(dA-dT) isolated from the DNA of the land crab, *Gecarcinus lateralis*); (b) 5 μg of DNA in alkaline CsCl gradient.

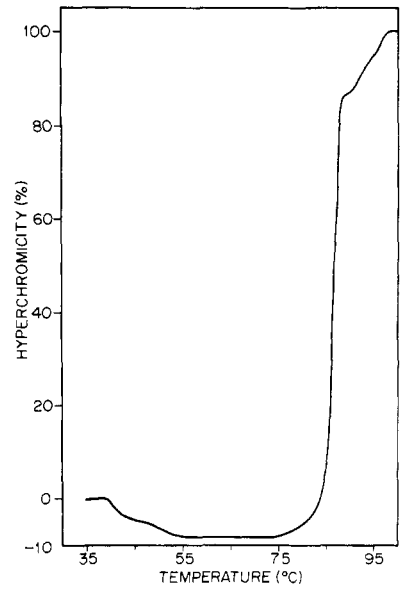

FIGURE 2: Thermal dissociation of satellite I: 43 μg of satellite I separated from satellite II by centrifugation in Cs_2SO_4 gradients containing Hg^{2+} and Ag^+ was heated in 0.12 M sodium phosphate buffer (pH 6.8) in stoppered 1-ml cuvets in a Gilford 2400 spectrophotometer programmed for a 0.5° temperature increase/min.

RNase and pancreatic RNase digestion. In both cases, the digestion was done in 3 μl containing 3 μg of enzyme plus 50 μg of carrier RNA in 10 mM Tris (pH 7.6). Digestion was for 30 min at 37°. The resulting oligonucleotides were separated by the two-dimensional fingerprinting method of Sanger *et al.* (1965). Electrophoresis on cellulose acetate strips, 3 × 85 cm (Wilson Diagnostics, Chicago), was in 6 M urea (pH 3.5) and pyridine acetate for 2.5 hr at 5000 V. Ionophoresis on DE81 paper, 40 × 85 cm (Reeve-Angel), was in 7% formic acid for 12 hr at 1100 V.

After location of the spots by autoradiography, radioactivity was quantitated on a low-background gas-flow counter and the oligonucleotides were eluted for further analysis as described by Sanger *et al.* (1965) and by Adams *et al.* (1969). Oligonucleotides produced by digestion with one RNase were treated with the other, and those resistent to enzyme treatment were hydrolyzed in alkali. Nucleotide sequences were deduced by analysis of nearest neighbor phosphate transfer from [α-^{32}P]NTP to product RNA (Billeter *et al.*, 1969; Bishop *et al.*, 1968; Lebowitz *et al.*, 1971).

Results

Single-Stranded DNA Templates. Three precautions were taken to ensure the highest possible purity of the single-stranded DNA templates. (1) The isopycnic twins were first separated from each other in the native state by centrifugation in $Hg^{2+} + Ag^+ - Cs_2SO_4$ gradients (Skinner and Beattie, 1973). The strands of satellite I were then separated by centrifugation in alkaline CsCl gradients (Figure 1). (2) The purity of isolated satellite I was monitored by thermal dissociation, a method that is far more sensitive than analytical centrifugation analysis for detecting small amounts of contaminating DNA. (3) After the final centrifugation in alkaline CsCl, in which the strands were separated from each other by 100 mg/cm^3, only the peak fractions were pooled to avoid recovering material in the region of intermediate density between the H and L strands. As can be seen (Figure 2), 90% of the purified satellite I showed a monophasic thermal dissociation; the remaining 10% may represent another satellite different from both satellite I and satellite II (see below).

Synthesis of RNA Products. The single-stranded DNAs of satellite I were effective templates for the synthesis of RNA products, the L strand (C-rich) being the better of the two. (It is possible that the solubility of the H strand was diminished because of its high guanine content.) Figure 3 shows the incorporation of label into acid-insoluble material in the presence of all four radioactive substrates.

With the exception of the product of the G-rich strand labeled with [α-^{32}P]GTP, which contained 6 × 10^3 cpm, the other products contained 2.4 × 10^4 cpm or more. This permitted localization of the degradation products after only a few hours of exposure of the autoradiograms. It also gave us statistically significant counts per minute of radioactivity for subsequent degradation and analyses.

Fingerprints. Two-dimensional paper electrophoresis separation of oligonucleotides (fingerprints) obtained after T_1RNase digestion of products using the two different DNA strands as templates are shown in Figure 4.

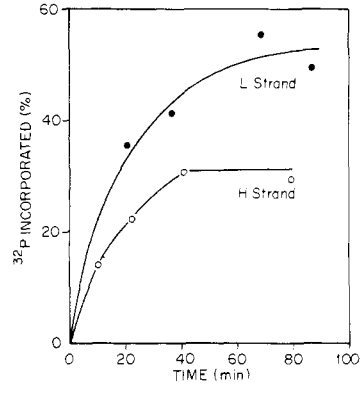

FIGURE 3: Incorporation of [α-^{32}P]NTP into RNA products transcribed from single strands of satellite I. Reaction mixture was as described in Materials and Methods. At times indicated, samples were removed and acid insoluble radioactivity in aliquots was counted as described. (●) L strand template; (○) H strand template.

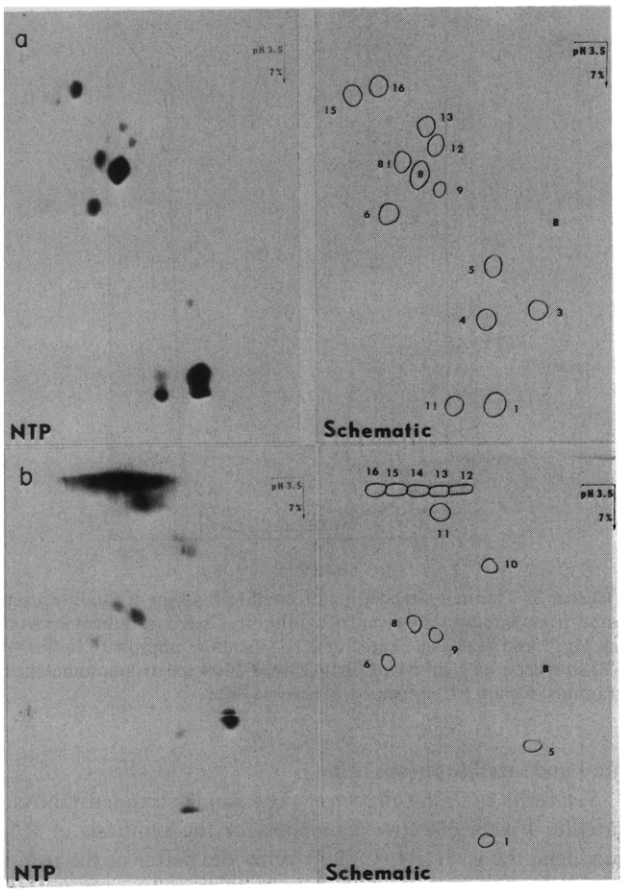

FIGURE 4: T_1 RNase fingerprints of RNA products synthesized *in vitro* by *E. coli* RNA polymerase using the L strand (a) and the H strand (b) as template DNAs and digested with T_1 RNase. All four ribonucleoside triphosphate precursors were ^{32}P-labeled in the α position when the product labeled "NTP" was synthesized. Fingerprints were as described in Materials and Methods. Also included is a schematic drawing showing the numbering system which was employed in Tables I and III, in which quantitations and nearest neighbor transfers are summarized. The position of the blue marker is indicated (B).

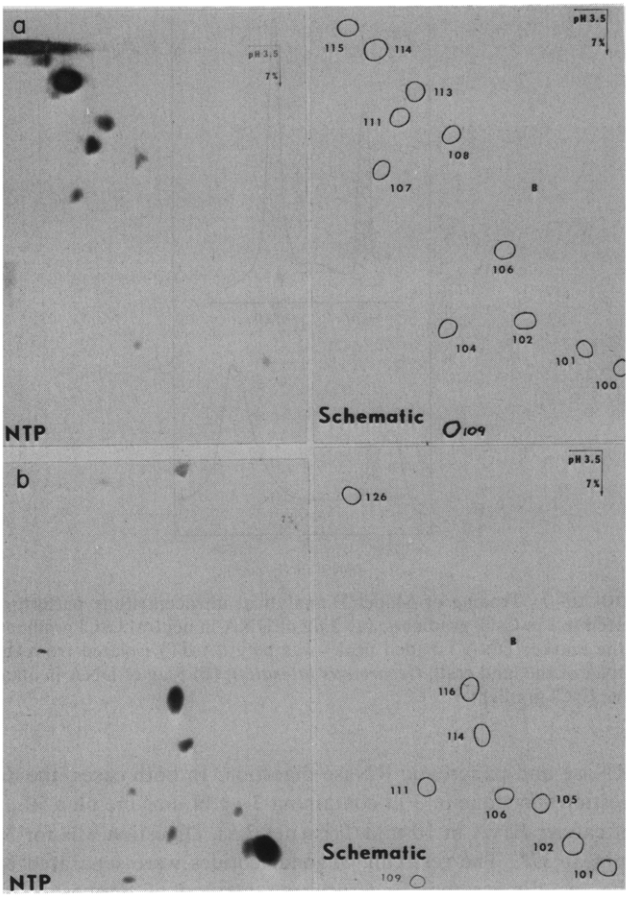

FIGURE 5: Fingerprints of RNA products using the L strand (a) and the H strand (b) as template DNAs and digested with pancreatic RNase. Details are given in the legend to Figure 4. Quantitations and nearest neighbor transfers are summarized in Tables II and IV.

These products were labeled either by including all four [α-^{32}P]NTPs in the reaction (Figure 4) or by including one labeled triphosphate and three cold triphosphates (fingerprints not shown). The numbering systems for the various oligonucleotides are shown in the schematic drawings in Figure 4. Pancreatic RNase fingerprints of the products of the two strands labeled with all four [α-^{32}P]NTPs and the numbering systems for the oligonucleotides are shown in Figure 5. Data from the RNA products obtained with each strand synthesized with one labeled triphosphate and three cold triphosphates are summarized in Tables I–IV.

A striking feature of the fingerprints is their simplicity. In every fingerprint one or two oligonucleotides are present in great excess over all others, both as assessed visually and by quantitation of radioactivity in each spot. This simplicity indicates that the RNA products (and hence the DNA templates) are composed of a short sequence repeated many times.

The patterns of oligonucleotides seen in the fingerprints are consistent with the overall base compositions of the respective template strands. For example, strand H (the template for the results presented in Figures 4b and 5b) contains 47% dG, approximately 25% each dA and dT, and virtually no dC (Skinner and Beattie, 1974). As expected, very little GTP incorporation was directed by this strand, so almost all oligonucleotides stayed on the origin of the second dimension in a T_1RNase fingerprint (Figure 4b). Similarly, very little C was found in the RNA synthesized from strand L (C-rich, G-poor) (Tables I and II).

Nucleotide Sequence of the Repeating Unit. The sequences of the major repeating unit can be deduced independently by analysis of the products synthesized from either strand. The repeating unit is a tetramer. Happily, the sequence deduced for the H-strand product, -C-C-U-A-, is complementary to the sequence of the other product, -U-A-G-G-.

The data used to obtain these sequences are summarized in Tables I–IV, where we indicate the oligonucleotides that were labeled when products were synthesized in the presence of the various [α-^{32}P]NTPs. In all cases we normalized the data to 1.0 mol of the most highly labeled spot in the fingerprint, based on the assumption that the entire sequence is primarily composed of a single repeating sequence.

A T_1 RNase digest of the uniformly labeled RNA product of strand L contains two major spots at the positions of Gp and U-A-Gp. This fact indicates that the sequence is -U-A-G-G-. In order to confirm this, we studied products made in the presence of individual labeled precursors. As seen in Table I, [α-^{32}P]GTP transferred its phosphate primarily to Ap and Gp in A-G- of U-A-G-; [α-^{32}P]ATP transferred its phosphate primarily to Up in U-A-G-. This is consistent only with the sequence proposed above. The same conclusion was reached by analysis of the pancreatic RNase fingerprints of the L-strand product (Figure 5a and Table II). In this case there was only one major oligonucleotide, which had the sequence -A-G-G-U. The 3′-phosphate of the U was derived from [α-^{32}P]ATP. The

TABLE I: Radioactive Oligonucleotides of RNA Products of Strand L Produced by Successive Treatment with T_1 and Pancreatic RNases.[a]

Oligo No.	α-^{32}P-Labeled Precursor: NTP a	NTP b	GTP a	GTP b	ATP a	ATP b	UTP a	UTP b	CTP a	CTP b	Summary a	Summary b
1	G	2.1	G	0.26	G	0.05	G	*1.00*	G	n.d.	G[U]	*1.00*
3	C; A-G	0.01	A*-G A*-G*	0.03	C	0.01			A-G	0.01	C-A-G[G] C-A-G[C]	0.01 0.01
4	A-G	0.05	A*-G*	0.10	A-G	n.d.	A-G	0.01			A-G[U]	0.01
5	A-A-G	0.01	A-A-G	0.02	A*-A-G	0.01					A-A-G[G]	0.01
6	U; G	0.16	U; G	0.21	G	n.d.	G	0.01			U-G[G] U-G[U]	0.20 0.01
8	U; A-G	1.00	A*-G*	*1.00*	U	1.00	A-G	0.06			U-A-G[G]	*1.00*
9			G; C; U				C	0.01	G; U	0.03	U-C-G[G] C-U-G[C]	0.01 0.02
12	U; A-A-G	0.02	A-A*-G A*-G*	0.02	U-A*-A-G	0.02	A-A-G	0.01	A-C; U	0.01	U-A-A-G[U]	0.02
14	A-U; A-G	0.02	A-G	0.04	A; U	0.01	A*-U	0.01			A-U-A-G[G]	0.02
15	U; G	0.02	U; G	0.005			U	0.03			U-U-G[G]	0.02
16	U; A-G	0.07	A*-G*	0.17	U	0.06	U; G; A*-U	0.09	U	n.d.	U-U-A-G[G]	0.06

[a] Ribonucleoside [α-^{32}P]triphosphates were used for labeling, either all four or each one separately as indicated. RNase digestions were performed on RNA products synthesized as described in Materials and Methods. The resulting oligonucleotides were separated by two-dimensional paper electrophoresis (Figures 4 and 5). Quantitation was done by cutting out the spot of paper containing the oligonucleotide and counting it in a Nuclear Chicago low background counter; counts per minute, normalized to 1.0 mol of the major spot (italicized in each case) are shown. The radioactive redigestion products (columns a) were obtained by subsequent pancreatic RNase digestion of each oligonucleotide originally produced by digestion of the product with T_1 RNase. The radioactivity incorporated (columns b) has been adjusted to represent relative numbers of labeled phosphates in the oligonucleotides. When there was a possible ambiguity about whether one or more phosphates in a pancreatic RNase digestion product were labeled, the oligonucleotide was eluted and further digested with alkali (0.2 N KOH, 37°; 20 hr). The resulting mononucleotides were separated by electrophoresis on Whatman No. 540 paper at pH 3.5 and labeled mononucleotides were located by autoradiography. Asterisks indicate the position of ^{32}P. The oligonucleotides of the repeat sequence are italicized in the summary column. n.d. indicates that although visually evident at the position of a known compound, the product was estimated to be present below the 0.1% level and its radioactivity was not determined. n.e. indicates that the product was present in trace amounts and, after counting, was not eluted from the spot on the first electropherogram and definitively identified. The radioactivity of the CTP-labeled product was normalized, assuming 0.01 mol of C-A-G[C].

double-spot appearance of this oligonucleotide resulted from separation of oligonucleotides terminating with a 3′-phosphate and a 2′:3′-cyclic phosphate (on the left), since the two oligonucleotides always had the same labeling pattern; for purposes of quantitation, the data of the two spots were combined.

The nucleotide sequence of the complementary RNA strand (made from the G-rich H-strand DNA) was determined only by the use of pancreatic RNase fingerprints (Figure 5). The reason for this is that, being very low in G, the bulk of the product was resistant to T_1RNase (oligonucleotides 12–16, Table III) and remained at the origin (Figure 4). The data in Table IV give nearest neighbor phosphate transfer frequencies for the various α-^{32}P-labeled ribonucleoside triphosphates. Again in this table, the data are normalized to represent 1.0 mol of the predominantly labeled oligonucleotide. From Table IV it can be seen that [α-^{32}P]CTP labels both Ap and Cp of -A-C-. [α-^{32}P]ATP labels Up and [α-^{32}P]UTP labels Cp. Thus the sequence of this RNA must be primarily -U-A-C-C-.

Minor Oligonucleotides: Variation and Contamination. The extent to which sequences other than the repeating tetramer are present can be evaluated by a study of some of the minor oligonucleotides seen in Figures 4 and 5. Analysis of these products is presented in Tables I–IV. Most of the minor oligonucleotides constitute only 1–5% of the RNA product. In principle they could be derived from contamination of the templates by a second repetitive DNA species, from contamination of the triphosphates used, or from actual variations in the primary structure of satellite I. Many of the results are most easily explained by low amounts of such contamination, but a small amount of sequence variation may be present.

Most of the minor sequences that are found in the products of strand L could have been derived from the major tetranucleotide repeat by one-step mutational events. The data are consistent with a 3–4% chance of a deletion of each nucleotide in each repeat, and about a 1% chance of an insertion of any nucleotide into any position in the repeat. Also, exchange of one nucleotide for any other appears to occur with a frequency of about 1%. It should be emphasized that not all of the "minor" spots are represented in the tables; only the ones that were seen consistently and that were present in reasonably abundant amounts (>1% of the major spots) are included.

Appearance of some of the minor oligonucleotides can be ascribed to impurities in the [α-^{32}P]NTPs used. For example, in an L-strand-templated reaction using [α-^{32}P]CTP, we observed a large amount of incorporation, even though the template is virtually lacking in dG. Analysis of the incorporated phosphate is consistent with the CTP being contaminated with 2–4% [α-^{32}P]UTP. This is not unexpected since CTP can easily

TABLE II: Radioactive Oligonucleotides of RNA Products of Strand L Produced by Successive Treatment with Pancreatic and T_1 RNases.[a]

Oligo. No.	α-^{32}P-Labeled Precursor										Summary	
	NTP		GTP		ATP		UTP		CTP			
	a	b	a	b	a	b	a	b	a	b	a	b
100			C	0.01	C	0.01	C	0.05	C	0.03	C[U]	0.05
101	A-C	0.03	A-C	0.01	A-C	0.02			A*-C*	0.05	A-C[C]	0.02
102	G; C	0.005					C	n.d.	G	0.04	G-C[U]	0.01
104	A-U	0.03	A-U	0.02	A-U	0.01	A*-U	0.05	A-U	0.02	A-U	0.03
105	A-G	0.01	A-G	0.07			C		A-G	0.03	A-G-C[U]	0.01
107	G; U	0.03			U	0.03	G	0.06	G; U	0.01	G-U [A]	0.03
108	A-G; U	0.02	A*-G	0.04			G	0.06	G	0.04	(A-G-G-C[A])	0.02
109	U	0.14	U	0.04	U	0.11	U	0.09	U	0.03	U	0.14
111	A-G; U	0.11	A*-C	0.05	U	0.07	A-G; U	0.09			A-G-U[A]	0.09
113	A-A-G; A-G; A-U; U	0.05	A-A*-G	0.03	U; A-U; A-G; A-A-G	0.03			A-A-G; U	n.d.	A-A-G-U[A]	0.03
114	A-G; G; U	*1.00*	A*-G*	*1.00*	U	*1.00*	G	*1.00*	G	0.04	A-G-G-U[A]	*1.00*
115	A-G; G; U	0.04	G; A*-G*	0.05	U	n.d.					A-G-G-G-U[A]	0.02

[a] Products were synthesized and analyzed, and data are presented as discussed in legend to Table I, except that each eluted oligonucleotide was digested with T_1 RNase and analyzed. Quantitation for CTP labeled product was based upon the assumption that there is 5% contamination of strand L by strand H.

TABLE III: Radioactive Oligonucleotides of RNA Products of Strand H Produced by Successive Treatment with T_1 and Pancreatic RNase.[a]

Oligo No.	α-^{32}P-Labeled Precursor										Summary	
	NTP		GTP		ATP		UTP		CTP			
	a	b	a	b	a	b	a	b	a	b	a	b
1	G	0.08	G	0.05	G	0.01	G	0.03	G	0.02	G[N]	0.05
5	C; A-G	0.10	A*-G	0.12	C	0.14			A-G	0.13	C-A-G[C]	0.12
7	U; G	0.03	U; G	0.03					n.e.	0.01	U-G[G] U-G[C]	0.03
8	C; U; G	0.03	U	0.04			C	0.04	G, U	0.03	C-U-G[C]	0.03
9	U; A-G	0.01	A*-G*	0.03	U	0.09	A-G; C; U	0.01	(AG; AC)	0.03	U-A-G[G]	0.01
10	A-C; C; U	0.01			U	0.03	C	0.01	A-C	0.03	C-C-U-A	0.03
11	A-C; C; U	0.03			U	0.01	C	0.04	A-C	0.03	C-C-U-A	0.03
12–16	A-C; C; U	*1.00*			U	*1.00*	C	*1.00*	A*-C*	*1.00*	A-C-C-U	*1.00*

[a] Products were synthesized and analyzed, and data are presented as discussed in legend to Table I. Quantitations were made relative to 1 mol of the repeating structure -A-C-C-U-[A]- on the origin of the second dimension of the fingerprints. The quantitation for GTP labeled RNA was made on the assumption that there are 0.12 mol of C-A-G-[C]/mol of A-C-C-U-[A].

be deaminated to form UTP. Because of the overall low incorporation of the authentic CTP, the UTP incorporation became significant.

A second source of contamination is in the DNA template strand. The data in Tables I–IV are consistent with there being approximately 1–3% cross-contamination between the H and L strands. For example, in Table II it can be seen that, with the L strand as template, [α-^{32}P]CTP labels a significant amount of A-Cp, [α-^{32}P]ATP labels Up, and [α-^{32}P]UTP labels a substantial amount of Cp. Each one of these transfers would require a double mutational event from the repeating tetramer, but the frequencies of the transfers, compared to the frequency of other sequence changes, are inconsistent with that explanation. Since these transfers are consistent with 3–5% contamination of strand L by strand H, we feel that DNA strand cross-contamination is the most reasonable explanation. Likewise, it appears that the H-strand preparation is contaminated by a low amount (3%) of the L strand, since the T_1 RNase and pancreatic RNase digestion products of strand H contain U-A-G(G) and A-G-G-U(A), respectively (Tables III and IV).

Another Satellite? In the course of this work we obtained evidence for the possible presence of another satellite DNA in our preparation. Because of the low level of this material, we cannot be certain whether it is both strands of a DNA with the

TABLE IV: Radioactive Oligonucleotides of RNA Products of Strand H Produced by Successive Treatment with Pancreatic and T_1 RNases.[a]

Oligo No.	α-^{32}P-Labeled Precursor: NTP a	NTP b	GTP a	GTP b	ATP a	ATP b	UTP a	UTP b	CTP a	CTP b	Summary a	Summary b
101	C	n.d.	C	0.02	C	0.12	C	*1.00*	C	0.37	C[U]	1.0
											C[C]	0.3
											C[A]	0.1
102	A-C	*1.00*	A-C	0.005	A-C	0.07	A-C	0.08	A*-C*	*1.00*	A-C[C]	1.0
105	A-A-C	0.04	n.e.	0.05	A*-A-C	0.07			A-A-C	0.08	A-A-C[C]	
106	G; C	0.04			n.e.	0.01	n.e.	0.02	G; C	0.035	G-C	0.04
109	U	*1.00*	U	0.05	U	*1.00*	U	0.09	U	0.05	U[A]	1.0
111	A-U	0.05			A-U	0.03	[A*-U]	0.06			A-U[A]	0.04
114	A-G; C	0.05	A*-G	0.09	C	0.06	C	0.02	A-G	0.09	A-G-C[A]	0.06
											A-G-C[U]	0.02
116	C; A; U[b]	0.22			U[b]	0.17	C[b]	0.03	A*-C*[b]	0.55	U-A-C-C[U] or C-U-A-C[C] or A-C-C-U[A]	0.02
126	n.e.	0.03	A*-G*	0.03	U	0.05	G	0.02			A-G-G-U[A]	0.03

[a] Products were synthesized and analyzed, and data are presented as discussed in the legend to Table I. The quantitation of the GTP labeled product was normalized on the assumption of 3% contamination of strand H by stand L although this level of contamination has no appreciable effect on the data. [b] The radioactivity in spot 116, which was underdigested U-A-C-Cp, on alkaline hydrolysis gave the indicated nucleotides. The radioactivity was distributed to the appropriate products before normalization of the data.

sequences d(-C-A-G-) and d(-C-T-G-) or a longer sequence with d(-C-T-G-C-T-G-C-A-G-C-A-G-C-A-G-C-A-G-) in one of the strands. In addition, we cannot be sure whether this is a new satellite or part of the one we are studying here. The sequence was determined from analysis of T_1 RNase oligonucleotides of the RNAs made from the H strand (Figure 4b, Table III). Since the template strand contained virtually no dC, it was very unusual that any oligonucleotides should be small enough to move from the origin in the second dimension of the electropherogram, yet six oligonucleotides did so. Two of these, numbers 10 and 11, contained no GMP and appeared to be T_1 RNase overdigestion products of the -A-C-C-U- product at the origin. Numbers 10 and 11 represent dimer and trimer segments of this sequence, respectively (T_1 RNase is reported to hydrolyze at Ap residues at 1–3% of the rate at Gp residues; Brownlee and Sanger, 1967). Two other oligonucleotides, Gp and -U-A-G-[G], correspond to the low level of strand L contamination in strand H.

The other two oligonucleotides in the fingerprint, spots 5 and 8, cannot be accounted for in any way except by the existence of a new DNA template. The sequence of the oligonucleotide in spot 5 is -C-A-G-[C] and that of spot 8 is -C-U-G-[C] (see Table III); they seem to be present at a ratio of approximately 3:1. Occurrence of these sequences by mutation from strand H seems quite unlikely, especially since that would require that two dCs (which are present in very low amounts in the template) be found separated by only two nucleotides. If these two oligonucleotides were part of the strand-H product, they should be found in the strand-L product, since they are complementary to each other. We do find them in the L product, although the levels are very low (<0.1%). We cannot decide unambiguously whether these came from variation or from the complement of this new sequence. -A-G-C-[A] and G-C-[U] are also present in pancreatic RNase digests, at about the 1% level.

Because quantitation consistently yields a 2.5- to 3-fold higher amount of spot 5 than spot 8, the sequence is probably composed of 3 mol of C-A-G[C] for each mole of C-U-G-[C]. Since pancreatic RNase digestion produces three A-G-C-[A] for each A-G-C-[U], the sequence may have two C-U-G-'s followed by four C-A-G-'s. Alternatively, the results are consistent with the trimeric sequences -C-A-G- and -C-U-G- on complementary strands, with 2–3 times as much of one strand as the other contaminating our preparation. There may also be variations on this sequence which occur at levels too low to be accurately assayed here.

Discussion

The points of enzymatic cleavage by pancreatic RNase and T_1 RNase of the RNA transcripts of the H and L strands are summarized in Table V, together with the DNA sequences which are derived from them.

Also included in the table are the mono- or oligonucleotides produced and the base to which the [α-^{32}P] is transferred. As noted, since T_1 RNase attacks next to guanine residues, the G-poor transcript of the H strand is insensitive to it and the products remain at the origin. The sequences of the two different DNA strands, which were obtained independently, are complementary, thereby confirming the generally accepted base-pairing hypothesis (Watson and Crick, 1953). The sequence agrees with the independently determined base composition of the template strands of about 25% each dA and dT and 50% either dG or dC (Skinner and Beattie, 1974).

From reassociation experiments (Southern, 1970; Hennig and Walker, 1970; Sutton and McCallum, 1971) and sequence analyses (Fry *et al.*, 1973) others have concluded that some satellite DNAs may not be perfectly repeating sequences. The results of our analyses with radioiodinated H strand show that, in this instance, minor amounts of other satellites may contaminate the preparations but that with a very sensitive technique, no strong case can be made for variation of sequences in satel-

TABLE V: Major RNA Products Transcribed from H and L Strands and Oligonucleotides Released from Them by Pancreatic or T_1 RNases.[a]

I. H strand

Template DNA: 3′ d(G-G-A-T-G-G-A-T-)$_n$5′
RNA product: 5′r(C-C-U-A-C-C-U-A-)$_n$3′
↑ ↑ ↑ ↑ ↑ ↑

	Pancreatic RNase Product Labeled			T_1 RNase Product Labeled		
Substrate	Identified by Position[b]	T_1 RNase Redigestion Products	Nucleotide Labeled	Identified by Position[b]	Pancreatic Redigestion Product	Nucleotide Labeled
[^{32}P]CTP	A-C-	A-C-	A-, C-	Origin[c]	A-C-	A, C
[^{32}P]UTP	C-	C-	C-	Origin[c]	C-	C-
[^{32}P]ATP	U-	U-	U-	Origin[c]	U-	U-

II. L strand

Template DNA: 3′ d(A-T-C-C-A-T-C-C-)$_n$5′
↓ ↓ ↓
RNA product: 5′ r(U-A-G-G-U-A-G-G-)$_n$3′
↑ ↑

	Pancreatic RNase Product Labeled			T_1 RNase Product Labeled		
Substrate	Identified by Position[b]	T_1 RNase Redigestion Products	Nucleotide Labeled	Identified by Position[b]	Pancreatic Redigestion Product	Nucleotide Labeled
[^{32}P]GTP	A-G-G-U	A-G-	A-, G-	U-A-G-	A-G	A-, G-
[^{32}P]ATP	A-G-G-U	U-	U-	U-A-G-	U-	U-
[^{32}P]UTP	A-G-G-U	G-	G-	G-	G-	G-

[a] ↑ point of hydrolysis by pancreatic RNase; ↓ point of hydrolysis by T_1 RNase. [b] Identity deduced from position on fingerprint. [c] All radioactivity remains at origin of second dimension of the fingerprint.

lite I. It appears then, that satellite I contains only three bases on each of its individual strands (Skinner and Beattie, 1974). In the course of this analysis, we discovered that our strand preparations have a low but significant amount of another sequence. The sequence of the RNA product of this second DNA was determined to contain -C-U-G- and -C-A-G-. From the experiments with radioiodinated H strand, it appears most likely that the sequence corresponds to a new satellite and is not an integral part of satellite I.

In the single-stranded DNAs used for these experiments, the level of contamination of strands with each other would be sufficient to account for the level of dC in strand H and dG in strand L. This consideration is all the more significant when one considers the fact that the contaminating strand is approximately 50% dC or dG. Thus, most of the dC and dG in strands H and L, respectively, may be accounted for by a low level of cross-contamination by the opposite strand. This would put the level of mutations to those nucleotides considerably below the levels for the other nucleotides. It remains to be seen whether such a bias against introduction of dC into strand H or dG into strand L is related to the function of satellite I.

The α satellite of the guinea pig is similar to crab satellite I in that it has interstrand bias in base composition, one strand containing 3% dG and the other 3% dC (Corneo *et al.*, 1968). We plan to reinvestigate whether the very small content of dC or dG in the opposite strands of the guinea pig satellite can be ascribed to contamination by other DNAs.

The major sequences of the guinea pig satellite strands, which have been determined to be hexamers and which make up approximately 50% of the satellite sequences (Southern, 1970), contain within them the same tetramers we report here for the hermit crab. Given the base composition of the H strand of hermit crab satellite I (1T:1A:2G), there can be only three sequences, allowing for circular permutation, and -T-A-G-G- occurs in only one of them. The sequence occurs in only two of the ten possible hexameric sequences that can be derived from the base composition (2T:1A:3G) of the guinea pig α satellite. This sequence also appears to be part of the most common repeat of the *Dipodomys ordii* HS-α satellite (W. Salser and K. Fry, in press). The fact remains that the sequence does occur in several cases that represent the majority of (dG + dC)-rich satellites that have been sequenced to date. Two factors beyond simple coincidence could account for the repeated recurrence of this sequence. First, the sequence may have been evolutionarily conserved for some as yet undetermined function although hermit crabs are not evolutionarily closely related to the other organisms listed. Second, and more mundane, the sequence results in such high density DNA in alkaline CsCl gradients that satellites containing it may have been "selected for" in the sense that they may be the easiest to isolate and analyze in various laboratories. Sequence determinations on more satellites will be required to settle this question.

The evolution of satellite DNAs has been the subject of several reports (Hennig and Walker, 1970; Mazrimas and Hatch, 1972; Sutton and McCallum, 1972; Graham and Skinner, 1973). In particular, Mazrimas and Hatch (1972) postulate that satellites confer greater genetic flexibility, a characteristic of more primitive species since they find that the DNA of more highly specialized kangaroo rats (*Dipodomys* sp.) contains significantly fewer satellite sequences than that of their primitive

counterparts. By contrast, the poly(dA-dT)-like satellite of numerous genera of Crustacea (Sueoka, 1961; Skinner, 1967; Skinner and Kerr, 1971) is distributed in varying amounts throughout the taxonomic suborder Brachyura (the true crabs). It comprises 6% of the DNA of the very primitive spider crab, *Libinia;* 10% (*Cancer irroratus*) to 30% (*C. borealis*) of the DNA of members of a genus of more highly advanced marine crabs and 18% of the DNA of the highly specialized land crab, *Gecarcinus*. Taxonomically sandwiched between the primitive and the highly specialized forms are the blue crab, *Callinectes,* the green crab, *Carcinus,* the fiddler crab, *Uca,* and others missing the poly(dA-dT) satellite although some species have other satellites (Sueoka, 1961; Skinner *et al.,* 1970; D. M. Skinner, unpublished results). Again, their amount shows no correlation with the species' taxonomic position. In brief, the quantity of satellites present in crustacean DNA shows no clear relationship with the degree of specialization or presumed genetic flexibility. Nor has a correlation been noted by others between the quantity of satellite DNA and the level of specialization of the organism (Hennig and Walker, 1970; Sutton and McCallum, 1972).

It is possible that the simple sequence satellite DNAs may confront physiological conditions at specific moments in the cell cycle that permit their destabilization. The satellite DNAs whose sequences have thus far been determined contain repeating dimeric (crabs; Swartz *et al.,* 1962; Gray and Skinner, 1974), trimeric and tetrameric (hermit crab, minor satellite, and satellite I), hexameric (guinea pig; Southern, 1970), heptameric (*Drosophila;* Gall and Atherton, 1974), and decameric (kangaroo rat; Fry *et al.,* 1973) units, each less than the minimum nucleotide number required to form a stable duplex (McConaughy and McCarthy, 1967). It is tempting to speculate that the destabilized satellites may pair with short stretches of homologous sequences on more complex DNA sequences and participate in chromosome recognition. A small number of the satellite-complementary sequences on the more complex DNA would allow for ease of destabilization and dissociation. The large number of such sequences on the satellite itself would allow for ease of recognition and possible longitudinal slippage which could be of functional significance in DNA dynamics. In addition to the satellites of simple sequence, there exist satellites of greater complexity that have nonbiased base composition between strands (Ingle *et al.,* 1973).

Acknowledgments

We thank Mr. John Dohrmann for crustacean tissues, Drs. Toby Mitchell and R. D. Wells for helpful discussions, and Dr. Wacław Szybalski for helpful discussions and the use of his laboratory.

References

Adams, J. M., Jeppesen, P. G. N., Sanger, F., and Barrell, B. C. (1969), *Nature (London) 223,* 1009.

Barrell, B. G. (1971), *Proced. Nucleic Acid Res. 2,* 751.

Billeter, M., Dahlberg, J. E., Goodman, H. M., Hindley, J., and Weissman, C. (1969), *Cold Spring Harbor Symp. Quant. Biol. 34,* 635.

Bishop, D. H. L., Mills, D. R., and Spiegelman, S. (1968), *Biochemistry 7,* 3744.

Blattner, F. R., and Dahlberg, J. E. (1972), *Nature (London), New Biol. 237,* 227.

Bollum, F. (1959), *J. Biol. Chem. 234,* 2733.

Brownlee, G. G., and Sanger, F. (1967), *J. Mol. Biol. 23,* 337.

Chamberlin, M., and Berg, P. (1962), *Proc. Nat. Acad. Sci. U. S. 48,* 81.

Corneo, G., Ginelli, E., Soave, C., and Bernardi, G. (1968), *Biochemistry 7,* 4373.

Fry, K., Poon, R., Whitcome, P., Idriss, J., Salser, W., Mazrimas, J., and Hatch, F. (1973), *Proc. Nat. Acad. Sci. U. S. 70,* 2642.

Gall, J. G., and Atherton, D. D. (1974), *J. Mol. Biol.* (in press).

Graham, D. E., and Skinner, D. M. (1973), *Chromosoma 40,* 135.

Gray, D., and Skinner, D. M. (1974), *Biopolymers 13,* 843.

Hennig, W., and Walker, P. M. B. (1970), *Nature (London) 225,* 915.

Ingle, J., Pearson, G. G., and Sinclair, J. (1973), *Nature (London), New Biol. 242,* 193.

Lebowitz, P., Weissman, S. M., and Radding, C. M. (1971), *J. Biol. Chem. 246,* 5120.

Loewen, P. C., and Khorana, H. G. (1973), *J. Biol. Chem. 248,* 3489.

Mazrimas, J. A., and Hatch, F. T. (1972), *Nature (London), New Biol. 240,* 102.

McConaughy, B. L., and McCarthy, B. J. (1967), *Biochim. Biophys. Acta 149,* 180.

Robertson, H. D., Barrell, B. G., Weith, H. L., and Donelson, J. E. (1973), *Nature (London), New Biol. 241,* 38.

Sanger, F., Brownlee, G. G., and Barrell, B. G. (1965), *J. Mol. Biol. 13,* 373.

Sanger, F., Donelson, J. E., Coulson, A. R., Lossel, H., and Fischer, D. (1973), *Proc. Nat. Acad. Sci. U. S. 70,* 1209.

Skinner, D. M. (1967), *Proc. Nat. Acad. Sci. U. S. 58,* 103.

Skinner, D. M., and Beattie, W. G. (1973), *Proc. Nat. Acad. Sci. U. S. 70,* 3108.

Skinner, D. M., and Beattie, W. G. (1974), *Biochemistry 13,* 3922.

Skinner, D. M., Beattie, W. G., Kerr, M. S., and Graham, D. E. (1970), *Nature (London), 227,* 837.

Skinner, D. M., and Kerr, M. S. (1971), *Biochemistry 10,* 1864.

Solymosy, F., Fedorcsak, I., Gulyas, A., Farkas, G. L., and Ehrenberg, L. (1968), *Eur. J. Biochem. 5,* 520.

Southern, E. M. (1970), *Nature (London) 227,* 794.

Sueoka, N. (1961), *J. Mol. Biol. 3,* 31.

Sutton, W. D., and McCallum, M. (1971), *Nature (London), New Biol. 232,* 83.

Sutton, W. D., and McCallum, M. (1972), *J. Mol. Biol. 71,* 633.

Swartz, M. N., Trautner, T. A., and Kornberg, A. (1962), *J. Biol. Chem. 237,* 1961.

Vinograd, J., Morris, J., Davidson, N., and Dove, W. F., Jr. (1963), *Proc. Nat. Acad. Sci. U. S. 49,* 12.

Watson, J. D., and Crick, F. H. C. (1953), *Nature (London) 171,* 737 (see also 964).

Ziff, E. B., Sedat, J. W., and Galibert, F. (1973), *Nature (London), New Biol. 241,* 34.

Zimmerman, S. B., and Sandeen, G. (1966), *Anal. Biochem. 14,* 269.