NOTES

Similarity between the *Myxococcus xanthus* and *Stigmatella aurantiaca* Reverse Transcriptase Genes Associated with Multicopy, Single-Stranded DNA

MEI-YIN HSU, CHUNYING XU, MASAYORI INOUYE,* AND SUMIKO INOUYE

Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey at Rutgers, Piscataway, New Jersey 08854

Received 14 November 1991/Accepted 28 January 1992

To determine the evolutional relationship of bacterial retroelements of Myxococcus xanthus and Stigmatellaaurantiaca, the nucleotide sequence of 3,060 bases encompassing msr, msd, and the upstream region of msd (downstream of msr) of S. aurantiaca DW4 was determined and compared with the same region from M. xanthus. An open reading frame was found 92 bases upstream of msd which encoded a polypeptide of 480 amino acid residues having 73% identity with the reverse transcriptase of M. xanthus. Together with high homologies in msr (86%) and msd (81%) regions, the present data indicate that the reverse transcriptase genes as well as the retrons of M. xanthus (retron-Mx162) and S. aurantiaca (retron-Sa163) were derived from a common progenitor retron which possibly existed before the two myxobacterial species diverged.

Multicopy single-stranded DNA (msDNA) was originally found in Myxococcus xanthus, a gram-negative soil bacterium, as a satellite DNA consisting of a single-stranded DNA of 162 bases (21). A highly homologous msDNA was subsequently found in Stigmatella aurantiaca DW4 (16), another myxobacterium, and its entire primary structure of the branched RNA linked msDNA was determined (2, 3). msDNAs from M. xanthus and S. aurantiaca are designated msDNA-Mx162 and msDNA-Sa163, respectively (9, 10). msDNA-Sa163 consists of a 163-base single-stranded DNA (msDNA) and a 76-base RNA (msdRNA). The 5' end of the DNA is attached to the 2'-OH group of the rG residue at position 19 of the RNA molecule by a 2',5'-phosphodiester linkage. The 3' ends of the DNA and RNA molecules are hybridized to each other by 8 bases. The proposed secondary structure of msDNA-Sa163 is shown in Fig. 1A (2, 3).

The structure of msDNA-Mx162 has also been determined and shows a high homology (>80%) to msDNA-Sa163 (6). msDNAs have been found in a variety of myxobacteria (4) and also in a minor population of natural isolates of *Escherichia coli* (for reviews, see references 9 and 10). In addition to msDNA-Mx162 and msDNA-Sa163, five other msDNAs have been identified: msDNA-Mx65 (5), msDNA-Ec67 (14), msDNA-Ec73 (18), msDNA-Ec86 (15), and msDNA-Ec107 (7). In all cases, msDNA is encoded from a single chromosomal locus, designated a retron, consisting of *msr*, *msd*, and the gene for reverse transcriptase (RT), which is essential for the biosynthesis of msDNA (for reviews, see references 9 and 10).

In this report, we determined the nucleotide sequence of the RT gene of S. aurantiaca and found that the S. aurantiaca RT consists of 480 amino acid residues and has 73% identity with the *M. xanthus* RT. Its evolutional implication will be discussed. The RT gene is the first gene sequenced in *S. aurantiaca*.

To determine the nucleotide sequence of the region upstream of msd (downstream of msr), the 3.6-kb BamHI fragment (from B^e to B^f; Fig. 1B) from pSTA1 was subcloned into the unique BamHI site of pUC9 (19) and the resulting plasmid was designated pSTA5 (Fig. 1B). In pSTA1, the 13.3-kb EcoRI fragment from S. aurantiaca (2) was cloned into the EcoRI site of pACYC184 (1). The 3.6-kb BamHI fragment was digested with XhoI, SmaI, and RsaI, and the fragments thus generated were subcloned into pUC9. The DNA sequence was determined by the chain termination method (17). The nucleotide sequence of 3,060 bases encompassing msr, msd, and the gene for RT was thus determined as shown in Fig. 2. A long open reading frame starting from an initiation codon (ATG) was found 92 bases upstream of msd. The open reading frame codes for a polypeptide of 480 amino acid residues, and a ribosomebinding sequence (AGG) is found 7 bases upstream of the initiation codon (residues 753 to 755). In Fig. 2, msd, the gene for msDNA, and msr, the gene for msdRNA, are boxed, and inverted repeat sequences consisting of two 33-base sequences immediately upstream of the branched G residue (residue 458) and downstream of the 5' end of msDNA are indicated by a1 and a2. These inverted repeat sequences have been shown to be essential in order to form a stem structure in the primary transcript placing the branched G at the end of the stem structure for the msDNA priming reaction (8).

When the nucleotide sequence of S. aurantiaca from residues 122 to 3060 is compared with the corresponding nucleotide sequence of M. xanthus (12), homologies of msr and msd are 86 and 81%, respectively (6). There are 18 base

^{*} Corresponding author.



substitutions, 16 base insertions, and 1 base deletion within the 92-base sequence between the 5' end of msDNA and the initiation codon (residues 763 to 765) in S. aurantiaca. Within the 1,440-base open reading frame (residues 763 to 2202), there are 281 base substitutions, 17 base insertions, and 32 base deletions. The homology of this region is 77% between S. aurantiaca and M. xanthus. The homology at the region of the 3'-end-encoding sequence (residues 2203 to 2205) drops to 50% as a result of many deletion mutations. As expected from the homology of the DNA sequence, the RT amino acid sequence deduced from the DNA sequence is significantly similar to that of RT-Mx162. M. xanthus has been shown to contain two independent retrons, retron-Mx162 and retron-Mx65, responsible for two different species of msDNA, msDNA-Mx162 (6) and msDNA-Mx65 (5). Each retron contains a unique region coding for its own RT, RT-Mx162 for retron-Mx162 (12) and RT-Mx65 for retron-Mx65 (11). Figure 3 shows the amino acid sequence of the S. aurantiaca RT together with those of RT-Mx162 and -Mx65. There are 351 identical amino acid residues out of 480 residues between RT-Sa163 and RT-Mx162 (73% identity). When the 239-residue sequence encompassing only the RT domain (reference 20 and see below) is compared with RT-Mx162, the identity increases to 80%. This high identity between the two RTs is unique among all the other known bacterial RTs, which do not share more than 40% identities. It is also interesting that the identity between RT-Sa163 and RT-Mx65 is only 35%, as is the identity between RT-Mx162 and RT-Mx65. These results indicate that retron-Sa163 and retron-Mx162 were possibly derived from a common progenitor retron which was most likely acquired before the two myxobacterial species diverged. Supporting this is the fact that all 20 independent, natural isolates of M. xanthus from various locations in the world contain retrons which have greater than 80% nucleotide sequence homologies to retron-Mx162 (13).

In contrast to *M. xanthus*, only 13% of *E. coli* natural isolates contain retrons, and these retrons show substantial diversities in their primary sequences and sizes of msDNAs, msdRNAs, and RTs (for reviews, see references 9 and 10).



FIG. 1. (A) Proposed structure of msDNA-Sa163. The sequence in the box is RNA. The branched rG is circled. (B) Restriction maps of pSTA1 and pSTA5. The locations and the orientations of *msr*, *msd*, and the RT gene are indicated with arrows. E, B, and X, *Eco*RI, *Bam*HI, and *Xho*I sites, respectively.

ccc	ACT	TCC	GGC	GCT	CGG	GCT	GCG	CGA	GGG	ccc	GTG	CGA	GCA	CAT	GAT	GGC	GCT	r 606 (6CT	60	606	GGG		TO	ATC	: cro	ACC	AAC	GCG	CTG	900 (cae (66C (SAC 6	NTG G	TG G V	TG A	JAG K	1500
CGT	CCA	GGT	CCG	GCA	CCG	ccc	CGA	GCA	GGA	AGC	ACT	GCG	TCA	GVC	ccc	cec	GGG	5 CCG (CCA	120	~				-			~	-	- -		ngg (-	-	STC 2	- 	GA C	TG C	CTG (xec	1560
GCT	CAT	CCG	CGC	GGA	GAC	GCG	CTC	CTA	CCT	GCG	GCG	CGA	GCC	стс	CGG	CCA	663	GCA	oot	180	V	D	M	K		F	7	P	8	V	T	W	P	R	V	K	G	L	L	R	
GTA	CCG	CGT	стс	ATT	GGA	TGG	GAA	AGT	GGT	GGC	GGT	GGA	GTG	00 0	ccc	cce	cca	000	GGA	240	770 1	601	. ee	CT			, YYC	CTG L	GCG	ACG	CTC (L CTG	GCG (CTG (L CIC	TCC /	LCC G	LAG C	20C (206 P	1620
GTC	ccc	ccc	GCA	GNA	GCT	CTG	GTT	CGA	CAC	GGA	CCC	CGA	GGC	606	CAC	CCC	CTI	CTT (CAC	300	- 	646	. GTC		• •		:	GGA		ACG	CTG	TAC	GTG	GCC 1	ANG	GGC ·	cer e	260 (scs (TG	1680
GCG	CCT	GGA	GTC	CTT	GGC	œc	GCY.	606	ATA	TAT	CGA	TGC	ecc	TOC	TTC	AAT	GAT	GTA (GAA	360	R	E	v	v	R	P	R	G	1	T	L	¥	A	A +30	,x	G	P	R	X	L	
CAC	GCA	AGC	CAC	000	GCC PN		GCC	ccc	CGG	CGG	***	GGC	AGG	TGC	GAC	OGA	ACG	ACA	GAC	420	200 P	CAA Q	: 660 6	GC A	c cci	T AC	TCT	CCG P	GCG Å	CTG L	ACG .	NAC -	GCG A	CTG '	TGC C	CTG (R R	L	D D	K	1740
ACT	-	GCG	AGC	GAC	CGA	6 56	YOC	TCC	CAA	GCC	ATC	-	CTC	AGC	GCC	t TCG	XGC	: GCC 7		480	CGG	CTO	: TC	: 60	6 CT	-	S ANG	CGG	CTG	GGC	TTC	ACG	TAC	ACG	ccc	TAT	606 6	DAT (EAC (CTG	1800
TGA	GCA	CGC	TOG	CTG	GCT	CTC	TCC	AGG	GTT	C96	TAG	TCS	GAG	TCG	COG	AGC	100	COC 1	ICT.		R	L	8		L	s	K	R	L	G	P	T	¥	T	R	<u>I</u>	<u>×</u>	<u>D</u>	_ <u>D</u>	L	
CCC CCC	CGC	TTG	CGC	CCC	TCT	CCA	TGA ACT	ATT TAA	6CA 667	06) CCT		TCT AGA	66C	CAA GTT	CCA	AGC	CTC	TTC T	TTG AAC	540	ACG	TT	TCI	C TG W	G CG	s cg R	G GCG A	AAG K	ANC K	TCC S	CGG R	Q	AAG K	E	L	P	L	A A	D	A	1840
GCT	CTC	TTC	CCT	ccs	GTG	AGT	ACC	TCT	ccc	SCC	606	GNG	CTG	AAC	CAA	CGA	600	-	cac	600	ccc	GT	. 60	а ст	350 G CT	с ст	6 609	CGG	GTG	ANG	GGT	CTG	CTG	GAG	ecc	ana	00T :	TTC /	ACG	CTG	1920
CCA	GAG	110	GGA	GGC	CAC	TCA	766	121	200	Câŝ	CCC	CTC	GAC	TTG	GTT	GCT	CC	TIG			P	v	*	L	L	L	A	R	۷	ĸ	G	v	L				6	*	т 		
CGT CCA	TTC	CCC	GGC	CCC	AGA TCT	OCT CCA	ACT TGA	CAC	CCC GCC	ACC	CCT	GAG	CCC 66C	CTC	ACC	CTA	660	TOC O	222 888	660	CAC H	c cc	5 G)	с м	KG AC		ig GT(V	; CAG Q	R	AAG K	GCC	AGC S	R	CAG Q	R	CTG V	T	606 G	L	V	1980
								1	1			_							_		GT	G			co	ж с		: GTI	r ccc	GGT	GCC	CGG	GTG	ccc	Cec	GAT	GTG	GTG	CGG	CGG	2040
¥66	TGA	GAA	GCT	OGT	GCC CGG	TTC	GGG	CCT		TCG	ACC	GCT	CGC	SCT CGA	000	TOG	000	TOC	CCT	720	v	X		: 1	1 1	, ,	: G	v	P	G	X	R	v	P	R	D	v	v	R	R	
		~	۶Ņ	Ň,	~~~			~	~~~	~		~~~	~~~								CT		C G	-					CAG	CAG	GGC G	X	CCC P	GGC	CCC P	ACC	ecc G	CAC I	ACG	CTG L	2100
	100				~			~~~				•••		N	T	Å	x	L	1		-						~ ~~			~	-		100	GAC	600	636		aac		600	2160
TCA	CAC	GTC	ccc	ccc	GCG	ccc	ccc	GTC	TCC	ecc	GAG	ccc	ccc	GCC	ccc	ACC	007		GAT	840	E				K (L A	Å	F	L	H	X	T	D	X	Ŧ	x	6	R	Å	
s	R	v	P	A	*	P	P	v	8	*		*	P	*	2	T	R	P	D		TT	c ca	16 C		66 C	NG 64	-	-	-	YYC	cec	CAG	ACC	ecc	TGA	cee	TCA	CTG	GTC	GTC	2220
GCC	GCG	AAG	CAG	CAC E	GCC	CGC R	R	GCC A	CAC	CAC	CAG I	GCG A	CTG	R	CTG	R	TOO	I ANG (GCC	900	7			2	R)	. 1	5 A	L	E	K	R	Q	T		-						
100	G 11	6 10	- 600	-		200	6 1 C	acc	100	ata.	-	C36	- C16	CTG.	-	600				960	co	6 60	C A	rc G	CAG	CG G	C GC	CGG	G ACG	GYC	CGT	CVC	ccc	CCA	GAT	CTC	CAT	600	ATG	CTG	2280
ĩ	T	T	A		6	Ŧ	D	A	W	v	R	Q	Q	L	v	À	K	6	v		GG	GA	T C	rg g	GC G	ST G	NA GA	A GA	C TTC	CCA	GCC	CAG	ACG	GAC	GAA	. ecc	CTG	ccic	ATC	CGA	2340
ece	606	-	G	ere	GAC	TTC	ere	TCS	crc	AGC	evc	A AG	cive	<u>مب</u> د	ece	ecc	TGO	-	ere	1020	TG	a c	rc c	rc g	cc o	66 G	SC GA	T CT	c ccc	GNG	GGG	CAC	COT	TCC	GAC	GTC	CGT	GCC	ATT	6CT	2400
				•				•													CA	c a	CA G	66 C	тсо	ce e	~ ~	A GO	C TTO	661	GTC	CGC	CGA	GAA	GAN	. ana	CAG	ccc	GGA	GAT	2460
K	K	K	A	E	A CCC	T	E	R	R	SCC A	Q Q	XAG K	R	L	A	W	E	A GOC	W	1080	GG	ic a	T C	NG G	TT C	rc o	66 CG	A 00	C ATC	: CTC	600	ecc	coo		CAA	ATC	CTT	CAG	CAG	CNG	2520
ANG	GCC	ACG	CAC	ATC	CAC	CAC	CTG	GGC	GTG	GGG	GTG	CAC	+10 TGG	o GAC	GAG	ecc	-		ccc	1140	60	T G	c c	TT G	GC G	GT G	CC AT	c cc	т ссі	. cci	CAG	CTC	: ccc	: ecc	GTG	: 080	GCT	GTC	ACT	CGC	2580
K	*	Ŧ	H	I	H	H	L	G	۷	G	۷	H	w	D	E	A	G	G	P			ic G	M G	TA G	a c	AT C	CC A7	T CA	6 CG	CT	GAT	GGC	: GCT			:	GCT	GTC	: cee	ACC	2640
CAC D	X	TIC	GAC	GTG V	GCC A	GCC G	CGC R	GAG E	GAG E	CGG R	GCC A	AAG K	GCC A	AAC W	GGC G	L	. CCC P	E GAG	666	1200	c	6 0	CA G	AT G	nc c	тт с	ac co	5 GA	c cer		: ATG	cca		. 600	: ATC	: 601	GAC		CMG	CTC	2700
TTG	GAC	TCG	GTC	GAG	GCG	CTG	GCC	***	GCG	CTG	GGC	ATC	TCC	GTG	TOO	cec	CTO	- csc	TOG	1260	CI	nc e	œ c	TC G	igg c	TG G	α α	a ca	A CT	. 660	CTC		- 100	: ccc	:	: 601	GAA	. GN	GAT	CTT	2760
L	D	8	v *1	50 E	A	L	*	K	*	L	G	I	8	v	8	R	L	R	w		~	хc	cc e	ag c	xac c	GT G	NG N		G C0	AT	. GAG					1 GCG	: CAG	CTO	: CIC	: OGA	2820
TTC	TCC	TTC	CAC	CGC R	CAG E	GTG V	evc D	ACG T	GGC G	ACG T	CAC	TAC Y	CNC Q	ACG T	TGG	E	I I	r ccc . P	AAG K	1320			GT G	cc 1	ст е	GA G	ca c				c cro	: 600	- 110	C GTC	AT:	r or	: 040	. CAG	; eN	GAA	2880
CGG	GAC	GGC	GGC	ллс	CGG	ACG	стс	ACC	ccc	ccs	ANG	ccc	GAG	стс	***		GT	CAG	CCC	1380	a		AC C	GA (TC C	sc c	90 30	T GI			. GAG	: GN	GT	r cre	c cri	c	: ecc	:		GAC	2940
R	D	G	G	ĸ	R	T	L	T	λ	P	K	R	E *2	L	x		۷	Q	R				er e	GT (ict e	GA C	346 C	x ca		r GC	: cci	-		A CAG	. CT		ATT		: eT:		3000
TGG	GTG	CTC	GCG	AAC	GTG	GTG	GAG	с <u>с</u> с	CTG	~~~	CTG	CAC	666	écc	GCG	CAC	:	TTC	ere	1440	c	.c e					TC G	x 4	.c .co	5 GC		: 00	-	a ca	a oc	T 00		- 601		2 ChC	3060
-	•	-		-	•	•	-	•	**		•	-	-	-		-		-	•		-																				

FIG. 2. Nucleotide sequence of 3,060 bases encompassing *msr*, *msd*, and the RT gene of *S. aurantiaca*. Base 1 corresponds to 468 bases upstream of the X^e site in Fig. 1B, and base 3060 corresponds to 727 bases downstream of the B^f site in Fig. 1B. The sequence from base 421 to base 720 which contains *msr* and *msd* is shown double stranded. The boxed regions of the upper strand (bases 440 to 540) and the lower strand (bases 508 to 670) correspond to the sequences of msdRNA and msDNA, respectively (3). The starting sites for msDNA and msdRNA are indicated by open arrows. The circled G at the position 458 is the branched rG of msdRNA linked to the 5' end of msDNA. Long solid arrows labeled with a1 and a2 represent inverted repeated sequences proposed to form the secondary structure in the primary RNA transcript which serves to prime msDNA synthesis (6). Amino acids are indicated by single letters. The YXDD sequence highly conserved among known RTs is boxed. X^e and B^f sites are indicated by arrows. Numbers on the right-hand side and numbers with asterisks represent numbers for bases and amino acids, respectively.

Sa163	RT	MTAKLESHVP	AAPPVSAEAP	APTRPDA	akqearrahh	EALRLRWKAI	EEAGGTDAWV	RQQLVAKGVA	ABEVDFESLS	77
Mx162	RT	MTARLDPFVP	AASPQAVPTP	ELTAPSSDAA	AKREARRLAH	EALLVRAKAI	DEAGGADDWV	QAQLVSKGLA	VEDLDFSSAS	80
Mx65	RT				MSWFDTTLSR	LKGLFSRPVT	RSTTGLDVPL	DAHGRPODVV	TETVSTSGPL	50
							• •		•	
Sa163	RT	DKQKAAWKEK	KKAEATERRA	QKRLAWEAWK	ATHIHHLGVG	VHWDEAGGPD	KFDVAGREER	AKANGLPEGL	DSVEALAKAL	157
Mx162	RT	EKDKKAWKEK	KKAEATERRA	LKRQAHEAWK	ATHVGHLGAG	VHWAEDRLAD	AFDVPHREER	ARANGLTE-L	DSAEALAKAL	159
Mx65	RT	KPQGHLRQVR	RDARLLPKGV	RRYTPGRKKW	MEAAEARRLF	SATLRTRNRN	LRDLLPDEAQ	LARYGLPV-W	RTEEDVAAAL	129
			•				• •			
				* * * *	* *	*				
Sa163	RT	GISVSRLRWF	SFHREVDTGT	HYQTWEIPKR	DGGKRTLTAP	KRELKAVQRW	VLANVVERLP	VHGAAHGFVA	GRSILTNALA	237
MX162	RT	GLSVSKLRWF	AFHREVDTAT	HYVSWTIPKR	DGSKRTITSP	KPELKAAORW	VLSNVVERLP	VHGAAHGFVA	GRSILTNALA	239
MX65	RT	GVSVGVLRHY	SIHRPRERVR	HYVTFAVPKR	SGGVRLLHAP	KRRLKALQRR	MLALLVSKLP	VSPQAHGFVP	GRSIKTGAAP	209
		• •• ••	••	•• •••	• • •		• • ••	• •••••		
	-							**	** ** *	
Salos	KT	HQGADVVVKV	DMKDFFPSVT	WPRVKGLLRK	GGLPENLATL	LALLSTEAPR	EVVRFRGETL	YVAKGPRALP	QGAPTSPALT	317
MX102	RT	HQGADVVVKV	DLKDFFPSVT	WRRVKGLLRK	GGLREGTSTL	LSLLSTEAPR	EAVOFRGKLL	HVAKGPRALP	QGAPTSPGIT	319
MAOS	KI.	NVGRRVVLKL	DLKDFFPSVT	FARVRGLLLA	LGIGIPVAAT	LAVLATESER	QPVELEGILF	HVPVGPRVCV	QGAPTSPALC	389
					•	• • •• •	• •	• •••		
				* * ***	•					
Sa163	RT	NALCIRIDER	LSALSKRIGE	TYTEVADOLT	FCUDDAFFCD	OFFIDIADAD	WATTTADUEC	WI PAPOPALU	DOWNDUODEC	207
Mx162	RT	NALCLKLDKR	LSALAKRIGE	TYTHYADDLT	PSWTEAKODE	DDD-T-ODDD	VALLARVAG	VUPAPOPDUN	PDATRYQANG	397
Mx65	RT	NAVLLRLDRR	LAGLARRYGY	TYTRYADOLT	FSCDDVT	ALE-D-VD	-ATAA-PV	VOFFCFFVN	PERTEVOPEC	356
						ALL-A-VA		-VQEEGFEVN	ABAINVQANG	350
					••		• •			
		*								
Sa163	RT	SRORVTGLVV	NEAPEGVPGA	RVPRDVVRRL	RAATHNREOG	KPGPTGRTLR	OLEGIANTI	MTDARKGRAP	I.RRI.RALEER	477
Mx162	RT	TRORVTGLVV	NAAGKDAPAA	RVPRDVVROL	RAAIHNRKKG	KPGREGESLE	OLKGMAAFTH	MTDPAKGRAF	LAOLTELEST	477
Mx65	RT	GAORVTGVTV	NTTL	GLSREERPRL	RAMLHOEARS	EDVEAHRAHL	DGLLAYVKML	NPEOAERLAR	RRKPRGT	427
			•	• •	•• •					
								•		

Sal63 RT QTA Mx162 RT ASAAPQAE

480 485

FIG. 3. Amino acid sequence alignment of RT-Sa163, RT-Mx162, and RT-Mx65. Amino acids conserved in both RT-Sa163 and RT-Mx162 are marked with solid circles. Amino acids conserved in all known RTs (20) are indicated with asterisks. The YXDD is boxed.

Therefore, retrons are considered to be acquired into the *E. coli* genome after this bacterial species was established.

From the sequence alignment of 82 bacterial and eukaryotic RTs, the RT structure can be divided into 7 subdomains and contains 17 conserved amino acid residues plus 25 functionally homologous residues (20). In this alignment, 14 out of 17 identical and 18 out of 25 functionally homologous residues are found in RT-Sa163. These identical and homologous residues are marked with asterisks in Fig. 3. The GenBank-EMBL accession number for the RT-Sa163 gene is M86352.

We thank Susan Eagle for the critical reading of the manuscript. This work was supported by a grant from the National Institute of General Medical Sciences (GM44012).

REFERENCES

- 1. Chang, A. C. Y., and S. N. Cohen. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J. Bacteriol. 134: 1141-1156.
- 2. Furuichi, T., A. Dhundale, M. Inouye, and S. Inouye. 1987. Branched RNA covalently linked to the 5' end of a singlestranded DNA in *Stigmatella aurantiaca*: structure of msDNA. Cell **48**:47–53.
- 3. Furuichi, T., S. Inouye, and M. Inouye. 1987. Biosynthesis and structure of stable branched RNA covalently linked to the 5' end of multicopy single-stranded DNA of *Stigmatella aurantiaca*. Cell **48**:55–62.
- Dhundale, A., T. Furuichi, S. Inouye, and M. Inouye. 1985. Distribution of multicopy single-stranded DNA among myxobacteria and related species. J. Bacteriol. 164:914–917.
- Dhundale, A., M. Inouye, and S. Inouye. 1988. A new species of multicopy single-stranded DNA from *Myxococcus xanthus* with conserved structural features. J. Biol. Chem. 263:9055–9058.
- 6. Dhundale, A., B. Lampson, T. Furuichi, M. Inouye, and S. Inouye. 1987. Structure of msDNA from *Myxococcus xanthus*: evidence for a long, self-annealing RNA precursor for the covalently linked, branched RNA. Cell 51:1105–1112.
- Herzer, P. J., T. Kawaguchi, S. Inouye, and M. Inouye. Direct integration of a retron (retron-Ec-107) into the *Escherichia coli* genome by replacing a palindromic 34-bp intergenic sequence. Mol. Microbiol., in press.

- 8. Hsu, M.-Y., M. Inouye, and S. Inouye. 1989. Structure requirements of the RNA precursor for the biosynthesis of the branched RNA-linked multicopy single-stranded DNA of Myxococcus Xanthus. J. Biol. Chem. 264:6214–6219.
- 9. Inouye, M., and S. Inouye. 1991. Retroelements in bacteria. Trends Biochem. Sci. 16:18-21.
- Inouye, M., and S. Inouye. 1991. msDNA and bacterial reverse transcriptase. Annu. Rev. Microbiol. 45:163–186.
- 11. Inouye, S., P. J. Herzer, and M. Inouye. 1990. Two independent retrons with highly diverse reverse transcriptase in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. USA 87:942–945.
- Inouye, S., M.-Y. Hsu, S. Eagle, and M. Inouye. 1988. Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. Cell 56:709–717.
- Lampson, B. C., M. Inouye, and S. Inouye. 1991. Survey of multicopy single-stranded DNAs and reverse transcriptase genes among natural isolates of *Myxococcus xanthus*. J. Bacteriol. 173:5363-5370.
- Lampson, B. C., J. Sun, M.-Y. Hsu, J. Vallejo-Ramirez, S. Inouye, and M. Inouye. 1989. Reverse transcriptase in a clinical strain of *E. coli*: its requirement for the production of branched RNA-linked msDNA. Science 243:1033–1038.
- Lim, D., and W. K. Maas. 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. Cell 56:891–904.
- Qualls, G. T., K. Stephens, and D. White. 1979. Morphogenetic movements and multicellular development in the fruiting mxyobacterium, Stigmatella aurantiaca. Dev. Biol. 66:270-274.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Sun, J., M. Inouye, and S. Inouye. 1991. Association of a retroelement with a P4-like cryptic prophage (retronphage φR73) integrated into the selenocystyl tRNA gene of *Esche*richia coli. J. Bacteriol. 173:4171-4181.
- Vieira, J., and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.
- Xiong, Y., and T. H. Eickbush. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9:3353-3362.
- Yee, T., T. Furuichi, S. Inouye, and M. Inouye. 1984. Multicopy single-stranded DNA isolated from a Gram-negative bacterium, *Myxococcus xanthus*. Cell 38:203–209.