A comparative analysis of the triloops in all high-resolution RNA structures reveals sequence-structure relationships

(Supplementary Materials)

Véronique Lisi and François Major

Institute for Research in Immunology and Cancer
Department of Computer Science and Operations Research
Université de Montréal
PO Box 6128, Downtown station
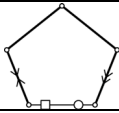*Montréal, Québec H3C 3J7*
CANADA

Correspondence to: François Major; Francois.Major@UMontreal.CA; +1(514)343-6752

Table S-I. PDB files. All PDB files where triloops were found.

1EFW	1GTR	1K9M	1N8R	1QRT	1U8D	1VQK	1Y39	1ZSE	2CSX	2HOL
1EUY	1GTS	1KD1	1NJI	1QRU	1U9S	1VQL	1YHQ	2A43	2CT8	2HOO
1EXD	1H3E	1L3D	1NTA	1QTQ	1VC6	1VQM	1YI2	2AZX	2CV0	2J00
1FFK	1HC8	1M90	1O0C	1QU2	1VQ4	1VQN	1YIJ	2B2D	2CV1	2J01
1FG0	1HR2	1MMS	1Q81	1QU3	1VQ5	1VQO	1YIT	2B57	2CV2	2J02
1FJG	1I9V	1MZP	1Q82	1QVG	1VQ6	1VQP	1YJ9	2BS0	2DR2	2J03
1G59	1J1U	1N32	1Q86	1S72	1VQ7	1XMQ	1YJN	2BS1	2FK6	437D
1GAX	1JJ2	1N77	1QA6	1SER	1VQ8	1Y26	1YJW	2BTE	2GDI	6MSF
1GID	1K8A	1N78	1QRS	1U0B	1VQ9	1Y27	1ZJW	2CKY	2GDI	

Table S-II Triloop instances. The 104 triloop specimens found in RNA-3A. Classes are named using Roman numbers (I to VIII) and shown on the left using the common topology of the structural graphs of its specimens. For each specimen, the structural subclass is indicated by a name that corresponds to its nucleotide interactions: L: phosphodiester Linkage with no combined stack or pair interaction; S: Stack; P: Pair. The four interactions along the backbone path constitute the first part of the name (i.e. I.LLLL means a subclass of class I defined by four phosphodiester Linkages with no other combined interactions). The dashes separate the backbone from the intra-loop interactions. The intra-loop interactions are listed after the first dash and are sorted by the first nucleotide involved in the interaction. The nucleotides are specified using the triloop numbering defined in Figure 1a. For example, subclass II.LLSL-S2-5 means a subclass of class II where nucleotides 1 and 2, 2 and 3, and 4 and 5 are simply Linked without any other interaction, nucleotides 3 and 4 are Stacked (and linked), and nucleotides 2 and 5 are involved in intra-loop Stacking. The subclasses with an asterisk after their names were previously reported in the literature. The numbers, underneath the subclass name and separated by a slash, are, respectively, the number of specimens and the number of triloop sites found in RNA-3A in this subclass. Each specimen is defined by a row indicating the number of sites (under Subclass), the structure number (Str) within the subclass, the RNA type (RNA), the PDB file identifier (PDB code), the PDB residue number of nucleotide 1 (A1 location), the sequence, the specific base pairing and stacking interaction types (Interactions), and the base pairing type of the flanking base pair with cis/trans (Flanking) and parallel/antiparallel notations (Para/anti). The letter 'n' in each row indicates that the specimen was previously reported (New). The rows that do not have an RNA, PDB code, and A1 location represent extrapolated sequences (see Materials & Methods) of the current structure. Each subclass is also represented by a structural graph (Diagram), where the specific or generalized interaction types are shown. The presence of specific base pairing or stacking types is indicated by corresponding symbols from the nomenclature (see Materials & Methods). However, when more than one base pairing or base stacking types are found, they are indicated, respectively, by black circled P or S. Grey circled S indicate stacking interactions at the limit of the automated annotation. These cases were examined visually and were determined as possibly stacking given small local motion.

Class	Subclass		Str	RNA	PDB code	A1 location	Sequence	Interactions			Flanking	Para/anti	New	Diagram
I	LLLL* 8/54													
		2	1	23S	2J01	A2749	AAGCA				H/W t	a		
		8	2	tRNA	1G59	B515	GCGGU				S/H c	p		
		5	3	5S	1Q86	B3022	GUUGC				H/W c	p		
		2	4	16S	1FJG	A1053	GCAUG				W/H c	a	n	
		1		RNase P	1U9S	A188	GCCCA				W/H t	a	n	
		32		23S	1FFK	0'335	UGACA				W/H t	a		
		2	5	16S	2J00	A201	CUUUG				W/W c	a		
		2		Intron I	1HR2	B235	AUCUU				W/W c	a		
	LLLS* 4/63													
		7	1	23S	1YIT	0'218	CGCGA	up			S/H t	a		
		21	2	5S	1YIT	9'22	GUUGC	up			S/W t	p	n	
		4	3	riboswitch	2B57	A47	UUUCU	up			W/H t	a		
		31		23S	1YJN	0'2482	GAUAA	up			W/H t	p	n	
	LLSL* 7/21													
		5	1	16S	1FJG	A64	GUGCG	up			W/H c	p	n	
		2		16S	1N32	A1053	GCAUG	up			W/H c	p	n	
		2		tRNA	2DR2	B54	UUCGA	up			W/H t	a	n	
		1		23S	1N8R	A2597	UUAAA	up			W/H t	a	n	
		1	2	16S	1FJG	A352	CAGCA	up			W/S c	p		
		2	3	23S	2J01	A1925	CUAAG	up			W/W c	p		
		8	4	23S	1VQ7	0'1186	CUAAG	in			W/W t	p	n	
	LLSS* 6/38													
		1	1	16S	1XMQ	A352	CAGCA	up	up		W/H c	p		
		2		16S	2J00	A618	CUCAA	up	up		W/Bh t	a		
		1	2	16S	1N32	A352	CAGCA	up	up		W/H c	p		
		3		16S	1FJG	A618	CUCAA	up	up		W/H t	a		
		1	3	23S	1VQN	0'1186	CUAAG	in	down		W/W t	p	n	
		30	4	23S	1JJ2	0'1813	UGACU	up	up		W/W c	a		
	LSLS 6/39													
		1	1	tRNA	1SER	T9	GCCCG	up	up		H/W t	a		
		1		tRNA	1B23	R9	AACAA	up	up		H/W t	a		
		14	2	23S	1N8R	A2033	GUCCC	up	in		W/Bs c	a		
		2		23S	2J01	A1992	GUCUC	up	in		W/Bs c	a		
		3	3	Virus	1L3D	A8	CACCG	up	up		W/H t	a		
		18	4	23S	1FFK	0'2033	GUCCC	up	in		W/W c	a		
	LSSS* 5/8													
		1	1	tRNA	1H3E	B9	UCCCG	up	up	up	H/W t	a		
		1		tRNA	1U0B	A9	AACAA	up	up	up	H/W t	a		
		1	2	Virus	2A43	A7	CACCG	up	up	up	W/H t	a		
		3	3	16S	1FJG	A934	CACAA	up	up	up	W/W t	p	n	
		2		Riboswitch	2GDI	X28	UGCGU	up	up	up	W/W t	p		
	PLLL 1/2													
		2	1	Intron I	1GID	A235	AUCUU	S/H t			W/W t	a		
	SLLL* 6/50													
		1	1	23S	1FFK	0'326	GAUAC	up			W/H c	p	n	
		2		16S	2J00	A1315	UGCAA	up			W/H t	a		
		1		tRNA	2CT8	C54	UUCGA	up			W/H t	a	n	
		1		tRNA	2J02	W54	UUCGA	in			W/H t	a		
		31	2	23S	1M90	A326	GAUAC	up			W/W c	a	n	
		14		23S	1N8R	A118	GAAUC	up			W/W c	a		

SLLS 1/32														
	32	1	23S	1FFK	0'2784	ACGCA	in	up		H/W	t	a		
SLSL* 24/272														
	21	1	tRNA	1EUY	B954	UUCGA	up	up		W/H	t	a		
	34		23S	1FFK	0'481	UGCAA	up	up		W/H	t	a	n	
	32		23S	1FFK	0'505	CGAAA	up	up		W/H	t	a	n	
	12		23S	1JJ2	0'624	UUUGA	up	up		W/H	t	a	n	
	36		23S	1FFK	0'1388	UGAGA	up	up		W/H	t	a	n	
	32		23S	1FFK	0'2597	UUAAA	up	up		W/H	t	a	n	
	32		23S	1FFK	0'313	UGGAA	up	up		W/H	t	a	n	
	9		tRNA	1FFY	T54	UUCAA	up	up		W/H	t	a		
	2		tRNA	1GAX	C953	UUCAA	up	up		W/H	t	a		
	3		16S	1FJG	A1177	GGAAG	up	up		W/H	t	a	n	
	1		16S	1FJG	A1315	UGCAA	up	up		W/H	t	a	n	
	5		16S	1FJG	A323	UGAGA	up	up		W/H	t	a	n	
	5		16S	1FJG	A956	UUAAU	up	up		W/W	t	p	n	
	1		RNase P	1U9S	A122	AGAGA	up	up		W/H	t	a		
	2		RNase P	1U9S	A170	UGAAA	up	up		W/H	t	a		
	8		tRNA	2AZX	C554	UUCGA	up	up		W/H	t	a		
	4		riboswitch	2H0J	A39	UGAGA	up	up		W/H	t	a		
	2		23S	2J01	A499	UGAAA	up	up		W/H	t	a		
	2		23S	2J01	A2562	UUAAA	up	up		W/H	t	a		
	2		23S	2J01	A306	UGGGA	up	up		W/H	t	a		
	2		23S	2J01	A475	UGAAA	up	up		W/H	t	a		
	1	2	tRNA	2J00	W54	UUCGA	in	up		W/H	t	a		
	23	3	23S	1YJW	A1186	CUAAG	up	in		W/W	t	p	n	
	1	4	23S	1HC8	D332	UUAAA	up	up		W/W	c	p		
SLSS* 4/8														
	1	1	23S	1Q81	A2597	UUAAA	up	up	up	W/H	c	a	n	
	3	2	23S	1HC8	C132	UUAAA	up	up	in	W/W	t	p		
	3		23S	1MMS	C1082	UUAAA	up	up	in	W/W	t	p		
	1	3	23S	1VQ8	0'1186	CUAAG	up	in	down	W/W	t	p	n	
II	**LLLL-S2-5** 1/11													
	11	1	23S	1FFK	0'842	CAAUA	out			W/S	c	a		
						UAAUA	out			W/S	c	a		
	LLSL-S2-5* 2/34													
	2	1	23S	2J01	A1752	CGCAG	up	out		W/W	c	a		
	32		23S	1FFK	0'1808	CGCAG	up	out		W/W	c	a	n	
						UGCAG	up	out		W/W	c	a		
	SLLL-S2-5 3/41													
	21	1	23S	1VQ8	0'842	CAAUA	up	out		W/S	c	a		
	2	2	23S	2J01	A319	CAGAG	up	out		W/W	c	a		
	18		23S	1VQ8	0'118	GAAUC	up	out		W/W	c	a		
						UAAUA	up	out		W/W	c	a		
						AGCUG	up	out		W/W	c	a		
						CGCUG	up	out		W/W	c	a		
						UGCUG	up	out		W/W	c	a		
	SSLL-P2-5 1/1													
	1	1	16S	1XMQ	A1028	CCCGG	up	up	W/W c	W/W	c	a		
						ACCGG	up	up	W/W c	W/W	c	a		
						GCCGG	up	up	W/W c	W/W	c	a		

III	LLLS-S2-4 4/17											
	3	1	16S	1FJG	A459	GACGA	up	up		S/H c	a	
	2		16S	2J02	A460	GACGA	up	up		S/H t	a	
	2		23S	2J01	A956	GAUAA	up	up		S/H t	a	
	10	2	23S	1FFK	0'218	CGCGA	up	out		S/H t	a	
						GAAGA	up	out		S/H t	a	
						GAUGA	up	out		S/H t	a	
	LLSS-P2-4* 1/32											
	32	1	23S	1FFK	0'1966	UUAAG	W/H c	up	in	W/W t	p	n
IV	LLLL-S1-3 5/30											
	17	1	23S	1FFK	0'671	AGUAU	up			Bs/H t	a	
	6	2	Aptamer	2BS0	S9	CUAAG	up			W/W c	a	
	1		Aptamer	6MSF	R6	GUCAC	up			W/W c	a	
	3		16S	1FJG	A201	CUUUG	up			W/W c	a	
	3		16S	1FJG	A838	GUCUC	up			W/W c	a	
						CAAAG	up			W/W c	a	
						CGAAG	up			W/W c	a	
						CAAUG	up			W/W c	a	
						CGAUG	up			W/W c	a	
						CUAUG	up			W/W c	a	
						GAAAC	up			W/W c	a	
						GCAAC	up			W/W c	a	
						CUAAC	up			W/W c	a	
						GCAGC	up			W/W c	a	
						GUAGC	up			W/W c	a	
						GAAUC	up			W/W c	a	
						GCAUC	up			W/W c	a	
						GUAUC	up			W/W c	a	
						GACAC	up			W/W c	a	
						GGCAC	up			W/W c	a	
						GACUC	up			W/W c	a	
						GGCUC	up			W/W c	a	
						GAGAC	up			W/W c	a	
						GCGAC	up			W/W c	a	
						GGGAC	up			W/W c	a	
						GUGAC	up			W/W c	a	
	LLLS-S1-3 1/2											
	2	1	23S	2J01	A247	GGCGA	up	up		S/H t	a	
						CACGA	up	up		S/H t	a	
						CGCGA	up	up		S/H t	a	
						CUCGA	up	up		S/H t	a	
						GACCA	up	up		S/H t	a	
						GACGA	up	up		S/H t	a	
						GAGAA	up	up		S/H t	a	
						GCGAA	up	up		S/H t	a	
						GGCCA	up	up		S/H t	a	
						GGGAA	up	up		S/H t	a	
						GUCCA	up	up		S/H t	a	
						GUCGA	up	up		S/H t	a	
						GUGAA	up	up		S/H t	a	
V	LLLL-S1-4* 1/32											
	32	1	23S	1FFK	0'1651	CCAUG	up			W/W c	a	n
						ACGGA	up			W/W c	a	

	Motif	n		Type	PDB	Residue	Seq								
	LSSS-P1-4 1/2														
		2	1	16S	2J00	A934	CACAA	up	up	up	W/H c	W/W t	p		
							CAACA	up	up	up	W/H c	W/W t	p		
							GCCGA	up	up	up	W/H c	W/W t	p		
VI	LLLL-S3-5 3/14														
		6	1	23S	1VQ8	0'138	UCGCG	out			H/S t		p		
							UCGGG	out			H/S t		p		
							UCGUG	out			H/S t		p		
		2	2	23S	2J01	A1798	UGCGA	out			S/Bs c		a		
							CCAAA	out			S/Bs c		a		
							CCAGA	out			S/Bs c		a		
							CCAUA	out			S/Bs c		a		
							CCGGA	out			S/Bs c		a		
							CCGUA	out			S/Bs c		a		
							UGCAA	out			S/Bs c		a		
		6	3	tRNA	1GTR	B33	UCUGA	up			S/H t		a		
							UCUAA	up			S/H t		a		
	LSLL-S3-5 5/84														
		1	1	Virus	1XJR	A22	GAGUA	up	up		S/H t		a		
		32		23S	1FFK	0'494	CAGAA	up	up		S/H t		a		
		30		23S	1KD1	A1707	GCGAA	up	up		S/H c		a		
		19		23S	1FFK	A1276	UCAUA	up	up		S/H c		a		
							GAGCA	up	up		S/H c		a		
							CAGUA	up	up		S/H c		a		
							GAAAA	up	up		S/H c		a		
							GAACA	up	up		S/H c		a		
							GAAGA	up	up		S/H c		a		
							GAAUA	up	up		S/H c		a		
							GAGAA	up	up		S/H c		a		
							GAGCA	up	up		S/H c		a		
							GCAAA	up	up		S/H c		a		
							GCACA	up	up		S/H c		a		
							GCAGA	up	up		S/H c		a		
							GCAUA	up	up		S/H c		a		
							GCGCA	up	up		S/H c		a		
							GCGUA	up	up		S/H c		a		
							UCAAA	up	up		S/H c		a		
							UCACA	up	up		S/H c		a		
							UCAGA	up	up		S/H c		a		
		2	2	23S	2J01	A642	GAACA	up	in		S/W c		p		
							GAAUA	up	in		S/W c		p		
							GCACA	up	in		S/W c		p		
							GCAUA	up	in		S/W c		p		
							GGACA	up	in		S/W c		p		
							GGAUA	up	in		S/W c		p		
	LSLS-S3-5* 2/33														
		1	1	23S	2J01	A476	GAAAA	up	down	up	S/H t		a		
		32		23S	1FFK	0'482	GCAAA	up	down	up	S/H t		a	n	
VII	LLLS-S1-3S2-4 1/4														
		4	1	23S	1VQ4	0'218	CGCGA	up	up	out	S/H t		a		
VIII	LLLL-P1-3S3-5 1/2														
		2	1	23S	1VQM	0'138	UCGCG	H/H c	out		H/S t		p		

Table S-III. Triloop modeling table. The 16 possible flanking base partner pairs are represented in a 4×4 matrix (first partner on the left and last partner at the top). For each partner pair, the possible flanking base pairing types are listed together with the observed and extrapolated (underlined) triplets pointing to specific triloop subclasses (class name followed by the subclass name). Three base partner pairs were neither observed nor extrapolated: CC, CU and UC, represented by a X in their corresponding matrix entries.

	A	C	G	U
A	H/W ACA I.LSLS H/W AGC I.LLLL H/W CGC I.SLLS W/H GAG I.SLSL W/W CGG V.LLLL-S1-4	✕	W/W CCG II.SSLL-P2-5 W/W GCU II.SLLL-S2-5	Bs/H GUA IV.LLLL-S1-3 W/W UCU I.LLLL W/W UCU I.PLLL
C	S/Bs CAA CAG CAU CGG CGU VI.LLLL-S3-5 S/H AGU VI.LSLL-S3-5 S/H ACG GCG UCG IV.LLLS-S1-3 S/H AGA VI.LSLL-S3-5 S/H GCG I.LLLS S/H GCG VII.LLLS-S1-3S2-4 S/H GCG III.LLLS-S2-4 W/Bh AGC UCA I.LLSS W/H GAA I.SLSL W/H UCA I.LLSS W/S AAU II.SLLL-S2-5 W/S AGC I.LLSL W/W AAU II.LLLL-S2-5 W/W AAC ACA V.LSSS-P1-4 W/W ACA I.LSSS	W/W UAA IV.LLLL-S1-3	W/H ACC I.LSLS W/H ACC I.LSSS W/W AAA AAU GAA GAU UAU UUU IV.LLLL-S1-3 W/W AGA II.SLLL-S2-5 W/W CAU V.LLLL-S1-4 W/W CCG II.SSLL-P2-5 W/W GCA II.LLSL-S2-5 W/W GCU II.SLLL-S2-5 W/W UAA I.LLSL W/W UUU I.LLLL W/W UAA I.LLSS W/W UAA I.SLSL	S/H CGG I.LSLS
G	S/H AAA AAC AAG AAU AGA AGC CAA CAC CAG CAU CGA CGC CGU VI.LSLL-S3-5 S/H ACG AAG III.LLLS-S2-4 S/H ACC ACG AGA CGA GCC GCG GGA UCC UCG UGA IV.LLLS-S1-3 S/H ACG AUA AUG III.LLLS-S2-4 S/H AGA LLLS-S1-3S2-4 S/H AGU VI.LSLL-S3-5 S/H CAA VI.LSLL-S3-5 S/W AAC AAU CAC CAU GAC GAU VI.LSLL-S3-5 W/H AGC I.LLSS W/H AUA I.LLLL W/H CCC I.LLLL W/W AAU II.SLLL-S2-5 W/W ACA IV.LLLL-S1-3 W/W CCG V.LSSS-P1-4	S/W UUG I.LLLS S/W UUG I.LLLS W/Bs UCC UCU I.LSLS W/H AUA I.SLLL W/W AAA AAU IV.LLLL-S1-3 W/W AAU AUA I.SLLL W/W AAU II.SLLL-S2-5 W/W ACU AGA CAA CAG CAU CGA GCA GCU GGA UAG UAU UCA UCU UGA IV.LLLL-S1-3 W/W UCC I.LSLS	H/W CCC I.LSLS W/H CAU I.LLLL W/H CAU UGC I.LLSL W/H GAA I.SLSL W/W CCG II.SSLL-P2-5	S/H CGG I.LLLL
U	S/Bs GCG GCA VI.LLLL-S3-5 S/H CAU CAA CAC CAG VI.LSLL-S3-5 S/H CUG CUA VI.LLLL-S3-5 S/H CAA I.LSSS W/H UAA I.LLLL W/H GAC I.LLLL W/H UAA UCG I.SLLL W/H GCA UCG I.SLSL W/W GAA GAG GCA GGA GGG UAA UCA UCG UUA UUG W/W AAU II.LLLL-S2-5 W/W AAU II.SLLL-S2-5	✕	H/S CGC CGG CGU VIII.LLLL-P1-3S3-5 H/S CGC CGG CGU VI.LLLL-S3-5 H/W CCC I.LSSS W/W GCA II.LLLS-S3-5 W/W GCU III.SLLL-S2-5 W/W UAA III.LLSS-P2-4	W/H UAA I.SLSL W/H UUC I.LLLS W/W GAC I.LLSS W/W GCG I.LSSS

Figure S-1. Secondary structure of the 23S rRNA subunit of *H. Marismortui*, showing the triloops found by (Lee et al., 2003) in various colors. The nine additional triloops we found are shown using cyan ellipses. The tertiary interactions of the new triloops are shown using orange lines for stacking interactions and pink lines for pairing interactions. The triloop nucleotides involved in the tertiary interactions are represented by the letter "R", followed by their nucleotide numbers in the triloop (1 to 5). The original Figure and authorization to update were kindly provided and given by Dr. Robin Gutell.