A Simple and Fast Approach for Predicting ^{1}H and ^{13}C Chemical Shifts: Toward Chemical Shift-Guided Simulations of RNA.

Aaron T. Frank*, Sean M. Law and Charles L. Brooks III*

Department of Chemistry and Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109-1055, United States

*Email: afrankz@umich.edu
brookscl@umich.edu

Table S1. Training Set. PDB and BMRB accession codes for NMR structures and chemical shifts for the RNA used to train LARMOR^{D}. Also included are the sample conditions under which spectra were assigned, the number of residues and brief description of each RNA.

PDBID	BMRB	T(K)	pH	#residues	Description
1KKA	5256	298	6.8	17	Anticodon Stem-Loop Of tRNA (Phe)
1L1W	5321	298	6.5	29	Helix 6 Of Signal Recognition Particle RNA
1LC6	5371	303	7.0	24	U6 RNA
1LDZ	4226	298	5.5	30	Lead-Dependent Ribozyme
1NC0	5655	303	6.5	24	U80g U6 ISL RNA
1OW9	5852	298	7.0	23	A Mimic Of The VS Ribozyme Hairpin Substrate
1PJY	5834	298	6.8	22	HIV-1 Frameshift Inducing Stem-Loop RNA
1R7W	6076	298	6.5	34	stem-loop IV domain of the enterovirus IRES
1R7Z	6077	298	6.5	34	stem-loop IV domain of the enterovirus IRES
1YSV	6485	303	6.0	27	Glur-B R/G RNA
2FDT	10018	293	6.0	36	RNA Hairpin Of Eel LINE
2GM0	7098	303	6.4	70	SL1 RNA in the HIV-1 DIS
2JXQ	15571	298	6.8	20	RNA Duplex
2JXS	15572	298	6.8	21	RNA Duplex Containing Single A-Bulge
2K3Z	15780	298	6.8	17	RNA Duplex with the C:G-A triple platform
2K41	15781	298	6.8	17	RNA Duplex with single U-Bulge
2KOC	5705	298	6.3	14	UUCG Tetraloop
2KYD	16980	303	6.5	32	RNA Duplex
2LBJ	17563	298	6.2	17	glycyl-tRNA Anticodon Stem-loop
2LBL	17565	298	6.2	17	Unmodified Glycyl-tRNA(UCC) anticodon stem-loop
2LDL	17671	303	5.5	27	HIV-1 ESS3
2LDT	17682	293	6.4	31	Helix-27 from *E. coli*16S rRNA
2LHP	17860	298	6.2	37	helix H1 of the Chimpanzee HAR1 RNA
2LI4	17877	293	6.8	32	RNA (32-MER)
2LK3	17972	298	7.0	24	U2/U6snRNA
2LP9	18239	298	7.0	16	Subgenomic Promoter Of Brome Mosaic Virus
2LPA	18240	298	7.0	15	Mutant sub-genomic promoter of Brome Mosaic Virus
2LU0	18503	300	6.7	49	κ-ζ region of S.cerevisiae group II intron ai5(γ)
2LUB	18515	298	6.2	37	Helix H1 Of The Human HAR1 RNA
2LV0	18549	298	6.8	24	Helix-35 Stem-Loop From E. Coli 23S rRNA
2RN1	11014	298	6.6	32	TAR-TAR*GA Aptamer Kissing Complex
2Y95	16714	298	6.0	14	AUCG Tetraloop
4A4S	18036	298	7.0	22	UCAC Tetraloop
4A4T	18034	298	7.0	22	UUAC Tetraloop
4A4U	18035	298	7.0	22	UGAC Tetraloop

Table S2. Testing Set. PDB and BMRB accession codes for NMR structures and chemical shifts for the RNA in the testing set. Also included are the sample conditions under which spectra were assigned, the number of residues and brief description of each RNA.

PDBID	BMRB	T(K)	pH	#residues	Description
1SCL	1SCL	303	7.6	29	SRL from 28S rRNA
1XHP	6320	283	6.4	32	U6 Extended ISL
1Z2J	6543	303	6.8	45	HIV-1 frameshift inducing RNA
1ZC5	6633	283	6.4	41	HIV-1 Frameshift RNA signal
2JWV	15538	298	7.0	29	anti-NF-κB RNA aptamer
2K63	15856	293	6.8	29	D3'-hairpin with EBS1 of group II intron Sc.ai5(γ)
2K64	15857	288	6.8	36	EBS1 + IBS1 of group II intron Sc.ai5(γ)
2K65	15858	288	6.7	14	EBS1/IBS1 of the group II intron Sc.ai5(γ)
2K66	15859	293	6.5	22	D3'-stem of the group II intron Sc.ai5(γ)
2LI4	17877	303	6.8	32	antiterminator hairpin from a Mg^{2+} riboswitch
2LPS	5962	303	6.7	34	domain 5 of the ai5(γ) group II intron
2LQZ	18336	298	6.5	27	RNA claw of the Φ29 DNA packaging motor
2LUN	18532	273	6.8	28	RNA Aptamer for *B. anthracis* Ribosomal Protein S8
2LX1	18656	274	6.1	22	RNA switch with 5'GAGU/3'UGAG internal loop
2M12	18838	277	6.4	23	ID3 stem loop of domain 1 in the ai5(γ) group II intron
2M21	18891	298	6.3	21	Tetrahymena telomerase RNA stem IV terminal loop
2M22	18892	298	6.3	23	helix II TBE from Tetrahymena telomerase RNA
2M23	18893	298	6.8	36	D3'-hairpin with EBS1 + IBS1 of group II intron Sc.ai5(γ)
2M24	18894	293	6.8	29	D3'-hairpin with EBS1 of group II intron Sc.ai5(γ)
2M8K	19260	283	6.3	48	*K. lactis* telomerase RNA pseudoknot
2MEQ	18975	298	7.3	19	Helix 69 from *E. coli* 23S rRNA
2MHI	19634	293	6.3	53	CR4/5 domain of medaka telomerase RNA
2MI0	19662	298	7.0	43	I-V kissing-loop interaction of the Neurospora VS ribozyme
2MIS	19692	273	6.5	26	Cations at the Active Site of the Neurospora VS Ribozyme
2QH2	7403	283	6.3	24	Telomerase RNA CR7
2QH3	7404	283	6.3	23	Human U64 H/ACA snoRNA
2QH4	7405	283	6.3	18	Human U85 C/D-H/ACA scaRNA

Table S3. Outliers in Datasets. Shown are entries in the training and testing sets identified as outliers using the 3σ-rule.

PDBID	Residue name	Residue	Nucleus	δ (ppm)	BMRB
1KKA	ADE	9	H5"	3.51	5256
1KKA	CYT	15	H6	8.46	5256
1KKA	CYT	16	H6	8.46	5256
1KKA	ADE	9	H4'	3.7	5256
1LC6	ADE	21	H4'	5.288	5371
1LDZ	ADE	25	C8	143	4226
1LDZ	CYT	11	H5"	4.59	4226
1PJY	CYT	8	C2'	93.761	5834
1R7W	CYT	23	C1'	95.481	6076
1R7W	ADE	15	H5"	4.64	6076
1R7Z	CYT	11	C5	141.052	6077
1R7Z	ADE	15	H5"	4.64	6077
1R7Z	CYT	23	H1'	3.769	6077
1SCL	CYT	8	H5	4.886	1SCL
1SCL	GUA	10	C3'	81.522	1SCL
1SCL	GUA	10	C1'	83.55	1SCL
1SCL	ADE	9	C2'	80.256	1SCL
1SCL	URA	11	C2'	72.149	1SCL
1SCL	ADE	20	H2'	5.126	1SCL
1SCL	URA	7	H3'	3.964	1SCL
1SCL	GUA	18	H3'	3.893	1SCL
1UUU	CYT	16	C4'	80.18	1UUU
1UUU	CYT	19	C3'	68.14	1UUU
1Z2J	CYT	21	H4'	3.613	6543
1ZC5	ADE	18	C2'	72.55	6633
2FDT	ADE	22	H1'	4.98	10018
2JWV	ADE	2	C5	68.677	15538
2JWV	URA	3	C5'	69.963	15538
2JWV	ADE	17	H3'	5.424	15538
2JXS	GUA	1	C5'	62.2	15572
2JYM	ADE	19	C4'	80.33	15745
2JYM	ADE	19	C5'	63.24	15745
2JYM	GUA	20	C4'	80.44	15745
2JYM	GUA	12	H4'	4	15745
2JYM	CYT	22	H3'	4.99	15745
2K41	GUA	9	C2'	78.02	15781
2K41	CYT	10	C5'	61.98	15781
2K63	CYT	26	C6	136	15856
2K64	ADE	10	C1'	87	15857
2K64	URA	9	C6	134.55	15857
2K64	CYT	29	H3'	3.91	15857
2K64	URA	11	C1'	88.07	15857
2K64	CYT	26	C5	94.42	15857
2K64	URA	24	C5	100.03	15857
2K64	ADE	6	C8	136.07	15857
2KOC	GUA	9	H3'	5.67	5705
2KOC	ADE	4	C6	159.6	5705
2KOC	CYT	8	H5"	2.73	5705

2KOC	GUA	10	C6	157	5705
2KOC	GUA	10	C2	162.5	5705
2KOC	GUA	10	C5	119.3	5705
2L1V	GUA	6	C8	133	17106
2L1V	CYT	10	H4'	4.84	17106
2L1V	URA	11	H2'	4.88	17106
2L1V	GUA	13	C8	133	17106
2L1V	ADE	16	C4'	87.27	17106
2L1V	CYT	20	H2'	5.01	17106
2L1V	URA	22	H4'	3.15	17106
2L1V	URA	22	H5"	3.77	17106
2L1V	URA	22	C1'	95.97	17106
2L1V	URA	24	H4'	4.82	17106
2L1V	URA	24	C4'	80.45	17106
2L1V	ADE	25	C3'	80.24	17106
2L1V	URA	26	H2'	2.8	17106
2L1V	URA	26	H6	6.69	17106
2L1V	CYT	33	C5'	97.47	17106
2LDL	ADE	7	C2	147.93	17671
2LDL	URA	8	C2'	82.27	17671
2LHP	URA	22	H5'	5.772	17860
2LHP	ADE	12	H5'	5.868	17860
2LI4	URA	19	H1'	3.816	17877
2LI4	URA	19	H1'	3.816	17877
2LI4	URA	19	H1'	3.816	17877
2LI4	URA	19	H1'	3.816	17877
2LI4	URA	19	H1'	3.816	17877
2LI4	URA	19	H1'	3.816	17877
2LI4	URA	19	H1'	3.816	17877
2LI4	URA	19	H1'	3.816	17877
2LK3	URA	9	H5"	4.637	17972
2LQZ	CYT	4	H5'	7.17	18336
2LQZ	URA	5	H3'	7.826	18336
2LQZ	CYT	16	C1'	86.483	18336
2LQZ	URA	15	C5	107.841	18336
2LQZ	URA	15	C6	145.633	18336
2LU0	GUA	27	H5'	5.771	18503
2LU0	GUA	19	H1'	3.689	18503
2LU0	GUA	32	C1'	94.54	18503
2LUB	GUA	21	H5"	4.678	18515
2LUB	GUA	2	H5'	5.771	18515
2LUN	URA	14	C2	155.84	18532
2LUN	GUA	8	C2	154.36	18532
2LUN	GUA	16	C6	162.35	18532
2LUN	CYT	15	H5	6.33	18532
2LUN	CYT	15	H5'	3.52	18532
2LUN	CYT	15	C2	160.81	18532
2LUN	CYT	27	C2	155.72	18532
2LUN	URA	20	C3'	70.4	18532
2LV0	GUA	1	C8	143.88	18549
2LV0	ADE	16	H3'	4.09	18549
2LV0	URA	8	C2	152.38	18549

2LV0	ADE	16	H2'	3.88	18549
2LV0	GUA	1	C3'	70.2	18549
2LV0	ADE	16	C3'	69.84	18549
2LX1	GUA	6	H2'	6.14	18656
2LX1	ADE	5	H5'	2.803	18656
2M21	URA	11	H5'	3.824	18891
2M22	ADE	6	H2	8.558	18892
2M24	GUA	8	C2'	72.449	18894
2M8K	GUA	1	H8	8.362	19260
2M8K	ADE	12	H8	8.673	19260
2M8K	CYT	30	C4'	85.99	19260
2M8K	GUA	36	C4'	86.07	19260
2M8K	CYT	22	C2'	71.69	19260
2M8K	ADE	24	H2	6.377	19260
2M8K	ADE	39	H8	7.149	19260
2M8K	URA	43	H5	4.898	19260
2M8K	ADE	38	C5'	69.17	19260
2MHI	GUA	21	H5'	3.818	19634
2MHI	GUA	22	H4'	4.687	19634
2MI0	URA	35	C3'	79.366	19662
2MI0	ADE	36	C1'	95.324	19662
2MI0	CYT	34	H6	6.951	19662
2MI0	URA	35	H5	6.171	19662
2MIS	URA	5	H1'	7.1	19692
2MIS	GUA	17	C5'	70.25	19692
2MIS	CYT	7	H1'	7.39	19692
2MIS	ADE	8	H1'	7.63	19692
2MIS	GUA	9	H1'	7.29	19692
2QH3	CYT	7	C3'	81.628	7404
2QH4	CYT	18	C6	146.4	7405
2QH4	ADE	8	C2	157.24	7405
2QH4	GUA	9	H5"	3.833	7405
2RN1	URA	7	C4'	94.42	11014
2RN1	GUA	4	H2'	3.256	11014
2RN1	GUA	11	H8	6.64	11014
2RN1	URA	6	H6	8.46	11014
2Y95	ADE	6	C5	120.7	16714
2Y95	GUA	12	C5	118.2	16714
2Y95	ADE	6	C6	157	16714

Table S4. Timing. LARMOR^{D} execution times for RNA in the testing set. Calculations were carried out on Intel Xeon 2.83 GHz processor.

PDBID	BMRB	#shifts	Time (s)
1SCL	--	319	0.31
1XHP	6320	425	0.43
1Z2J	6543	337	0.50
1ZC5	6633	585	0.71
2JWV	15538	408	0.37
2K63	15856	261	0.26
2K64	15857	312	0.34
2K65	15858	128	0.08
2K66	15859	203	0.16
2LI4	17877	231	0.30
2LPS	5962	309	0.37
2LQZ	18336	246	0.23
2LUN	18532	381	0.33
2LX1	18656	141	0.11
2M12	18838	269	0.23
2M21	18891	304	0.23
2M22	18892	325	0.25
2M23	18893	320	0.37
2M24	18894	286	0.27
2M8K	19260	517	0.74
2MEQ	18975	275	0.17
2MHI	19634	722	1.15
2MI0	19662	372	0.52
2MIS	19692	367	0.30
2QH2	7403	291	0.23
2QH3	7404	289	0.22
2QH4	7405	252	0.15

Table S5. Comparing accuracy of LARMOR^{D}, SHIFTS and NUCHEMICS: The mean absolute error (MAE) and Pearson correlation coefficient (*R*) between measured and LARMOR^{D}, SHIFTS and NUCHEMICS predicted ^{1}H chemical shifts. Results correspond to those obtained using the testing set (**Table S2**).

	MAE (ppm)			*R*		
nucleus	LARMOR^{D}	SHIFTS	NUCHEMICS	LARMOR^{D}	SHIFTS	NUCHEMICS
H1’	0.18	0.32	0.25	0.42	0.20	0.33
H2’	0.14	--	0.26	0.66	--	0.33
H3’	0.12	0.30	0.16	0.46	0.38	0.40
H4’	0.09	0.53	0.11	0.48	0.16	0.48
H5’	0.17	0.65	0.16	0.17	0.24	0.37
H5’’	0.13	0.34	0.15	0.16	0.18	0.26
H2	0.24	0.35	0.34	0.64	0.71	0.71
H5	0.15	0.27	0.26	0.69	0.58	0.54
H6	0.11	0.28	0.21	0.61	0.53	0.54
H8	0.18	0.32	0.23	0.77	0.44	0.71
mean	0.15	0.37	0.21	0.51	0.38	0.46

Table S6. Reference Chemical Shifts: Listed are the reference chemical shifts used by LARMOR^{D}.

Residue	Nucleus	Reference (ppm)
GUA	H5"	3.58
GUA	C3'	73.90
GUA	C1'	93.10
GUA	C5'	66.30
GUA	C4'	83.10
GUA	C2'	76.30
GUA	H2'	4.02
GUA	H1'	5.05
GUA	H4'	4.00
GUA	H8	7.69
GUA	H5'	3.78
GUA	H3'	4.14
GUA	C8	137.90
ADE	H4'	4.00
ADE	H8	8.60
ADE	H5'	3.78
ADE	H3'	4.14
ADE	H2	8.68
ADE	C8	141.20
ADE	C3'	73.90
ADE	C1'	93.10
ADE	C5'	66.30
ADE	H5"	3.58
ADE	C4'	83.10
ADE	C2'	76.30
ADE	C2	154.30
ADE	H2'	4.02
ADE	H1'	5.03
URA	H4'	4.00
URA	H5'	3.78
URA	H3'	4.14
URA	C3'	73.90
URA	C1'	93.10
URA	H5	5.95
URA	C6	142.00
URA	C5'	66.30
URA	H5"	3.58
URA	C4'	83.10
URA	C2'	76.30
URA	H6	7.85
URA	H2'	4.02
URA	C5	103.50
URA	H1'	5.60
CYT	H4'	4.00
CYT	H5'	3.78
CYT	H3'	4.14
CYT	H6	7.80
CYT	C3'	73.90
CYT	C1'	93.10
CYT	H5	6.20

CYT	C5'	66.30
CYT	H5"	3.58
CYT	C4'	83.10
CYT	C2'	76.30
CYT	H2'	4.02
CYT	C6	142.20
CYT	C5	98.20
CYT	H1'	5.28

Table S7. Imino ^{1}H and ^{15}N Chemical Shifts: In addition to predictors for non-exchangeable ^{1}H and protonated ^{13}C, predictors were also generated for imino ^{1}H and ^{15}N nuclei. However, *because of the paucity of imino chemical shifts in the training database, and uncertainties about the referencing of imino ^{15}N, these predictors should be considered "experimental" and "tentative"*. The table below summarizes the results of the application of the imino predictors to the testing set.

nucleus	MAE (ppm)	R
H1	0.37	0.85
H3	0.46	0.77
N1	1.30	0.29
N3	1.34	0.54

Table S8. Comparing the accuracy of measured and predicted chemical for the U6 ISL RNA: The mean absolute error (MAE) and Pearson correlation coefficient (*R*) between measured and LARMOR^{D} predicted ^{1}H and ^{13}C chemical shifts. Shown are the mean MAE and *R* for ^{1}H and ^{13}C chemical shifts, respectively. Analysis was restricted to chemical shifts for the region spanning residues 65-69 and 77-82. Results are shown for the pH 5.7 NMR structure (1SYZ), the pH 7.0 NMR structure (1SY4), the five CS-MD structures in Fig. 5D and the five MD structure in Fig. 5E.

	MAE (ppm)		*R*	
	^{1}H	^{13}C	^{1}H	^{13}C
1SYZ	0.19	1.19	0.29	0.74
1SY4	0.15	0.94	0.50	0.79
CS-MD1	0.12	0.52	0.69	0.90
CS-MD2	0.14	0.63	0.69	0.85
CS-MD3	0.14	0.71	0.63	0.86
CS-MD4	0.13	0.73	0.60	0.82
CS-MD5	0.13	0.64	0.64	0.88
MD1	0.20	0.88	0.39	0.58
MD2	0.20	0.96	0.35	0.66
MD3	0.19	0.98	0.27	0.78
MD4	0.16	0.87	0.40	0.83
MD5	0.18	0.97	0.29	0.63

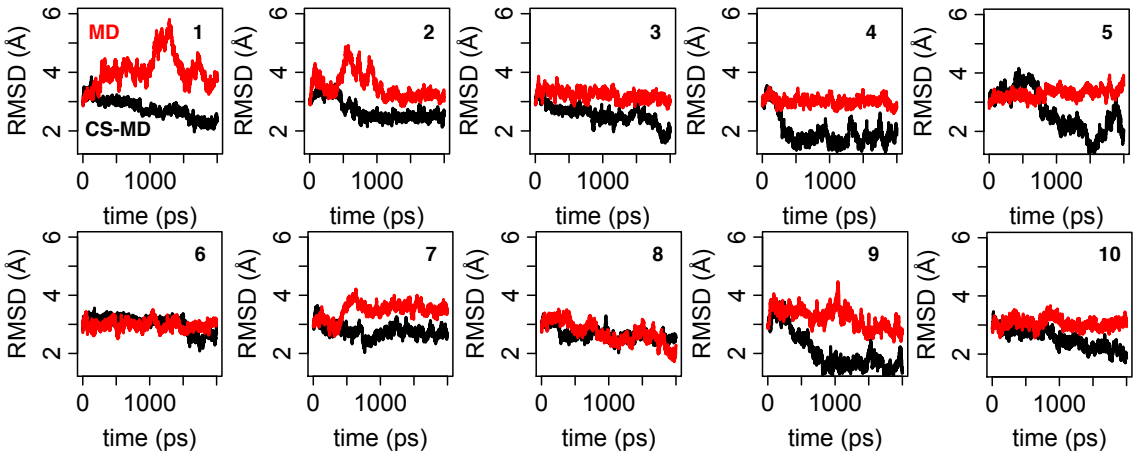

Figure S1. CS-MD and MD Simulations of U6 ISL RNA: Shown are the RMSD time series for the set 10 independent CS-MD (black) and MD (red) simulations of the U6 ISL RNA. RMSD were calculated relative to the first model of pH 7.0 NMR bundle (PDB: 1SY4) and for region spanning residues 65-69 and 77-82.

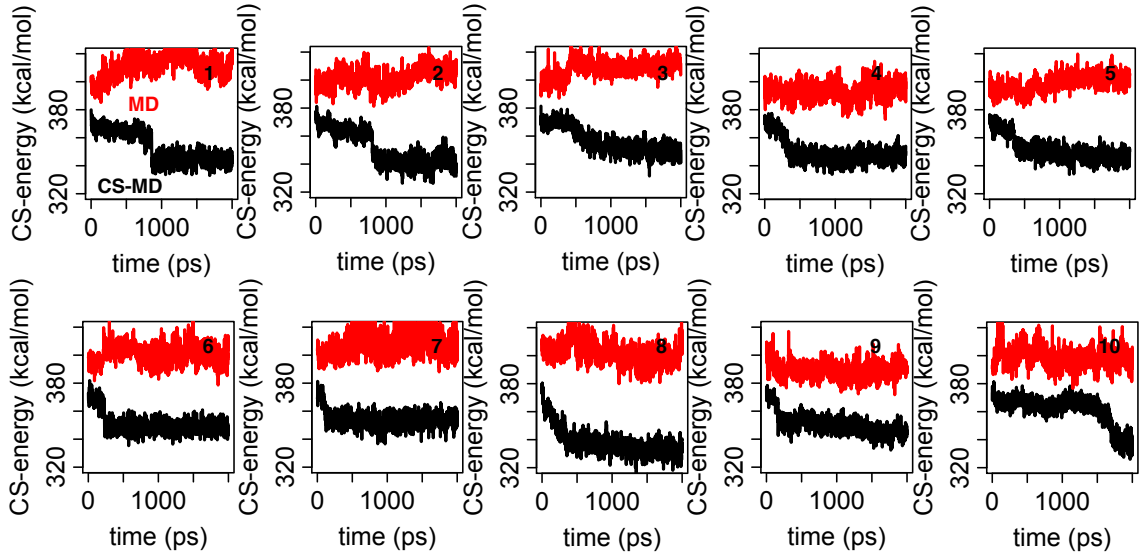

Figure S2. CS-MD and MD Simulations of U6 ISL RNA: Shown are the effective chemical shift error time series for the set 10 independent CS-MD (black) and MD (red) simulations of the U6 ISL RNA.