Supplementary Material

ID	Description	# Nucl.
157D	RNA duplex	24
1CSL	HIV-1 rev binding site	28
1D4R	Human SRP helix 6	58
1DQF	5S rRNA	19
1DUH	4.5s RNA	90
1DUQ	HIV-1 rev binding site	26
1F1T	RNA duplex	38
1F27	RNA pseudoknot	30
1FIR	HIV reverse-transcription primer tRNA(Lys,3)	76
1G2J	RNA duplex containing phenyl-ribonucleotides	16
1I9V	Phenylalanine tRNA	76
1I9X	snRNA	26
1J9H	RNA nonamer	18
1JZV	Bulged RNA from the SL2 stem-loop	17
1K9W	HIV-1 RNA dimerization initiation site	46
1KD5	RNA duplex	22
1KFO	RNA duplex	38
1KH6	RNA tertiary domain	53
1KXK	Ai5gamma group II self-splicing intron	70
1L2X	Viral (beet western yellow virus) rna pseudoknot	28
1L3Z	RNA heptamer	14
1MHK	26-nucleotide RNA containing a hook-turn	52
1MME	Hammerhead ribozyme	82
1MSY	Sarcin/Ricin domain from 23s RNA	27
1NBS	Ribonuclease P RNA	155
1NUJ	Leadzyme	24
1P79	Bulged RNA tetraplex	20
1Q93	Sarcin/Ricin domain mutant from 28s RNA	27
1QBP	RNA duplex	30
1SA9	RNA octamer	16
1SDR	Shine-Dalgarno region of 16S rRNA	24
1T0D	Ribosomal decoding site	33
1T0E	Ribosomal decoding site	35
1U9S	Ribonuclease P RNA	159
1X8W	Tetrahymena ribozyme	247
1X9C	Hairpin Ribozyme	61
1X9K	Hairpin Ribozyme	62
1XJR	SARS virus genome stem-loop II motif (s2m)	47
1Y0Q	Group I ribozyme	248
1YFG	Yeast initiator tRNA	75
1YKQ	Diels-Alder Ribozyme	49
1YZD	RNA duplex	32
1Z43	SRP19 RNA	101
1Z58	23s ARN Ribosomal	2877

ID	Description	# Nucl.
205D	RNA duplex	24
255D	RNA duplex	24
259D	RNA octamer	16
280D	RNA duplex	24
2A0P	RNA octamer	16
2A2E	Ribonuclease P RNA	334
2A64	Ribonuclease P RNA	414
2AO5	RNA duplex	20
2B8R	HIV-1 RNA dimerization initiation site	46
2G32	RNA octamer	16
2G3S	RNA octamer	16
2G91	RNA nonamer	17
2H0S	glmS ribozyme	145
2IL9	Ribosomal Binding Domain of the IRES RNA	282
2NOK	RNA internal ribosome entry site (IRES) domain	44
2OE6	Ribosomal decoding A site	33
2TRA	Yeast aspartic acid tRNA	73
333D	RNA octamer	16
353D	5S rRNA	23
357D	5S rRNA	61
361D	5S rRNA	20
377D	Alternating A-RNA hexamer	12
387D	RNA pseudoknot with 3d domain swapping	52
397D	HIV-1 trans-activation response region RNA stem	27
402D	RNA octamer	16
405D	RNA duplex	32
406D	RNA duplex	34
409D	RNA octamer	16
413D	RNA duplex	26
433D	RNA duplex	28
434D	acceptor stem of tRNA(Ala) from Escherichia coli	14
438D	RNA nonamer	18
472D	Octamer RNA with G.G/U.U tandem wobble base pairs	16

Table 1: **Reference set.** The *reference set* contains 77 molecules from [Bernauer et al., 2011]. The structure of theses molecules was used to compute the length of edges, the set strategies for 2-way junctions and the scoring functions. For scoring functions, the molecule 1Z58 was not used because it is very large compared to the other studied molecules.

ID	Description	# Nucl.	# Players
1E8O	7SL RNA	50	8
1MZP	Fragment of 23S rRNA	55	8
4FE5	Guanine riboswitch aptamer domain	68	14
2DU3	O-phosphoseryl-tRNA synthetase	71	13
4QJH	Twister Ribozyme	74	15
3Q3Z	c-di-GMP-II riboswitch	77	9
1P5O	77-MER	77	19
2HOJ	Thi-box riboswitch	79	15
4FRG	Cobalamin riboswitch aptamer domain	84	17
4TS0	Spinach RNA aptamer	89	21
2GIS	SAM-I riboswitch	94	21
1LNG	7S.S SRP RNA	97	16
4WFL	Bacterial SRP Alu domain	107	18
1C2X	5S RIBOSOMAL RNA	120	17
4QK8	C-di-AMP riboswitch	124	20
1MFQ	7S RNA of human SRP	127	24
1GID	P4-P6 RNA Ribozyme Domain	158	27
3D0U	Lysine Riboswitch RNA	161	32
2QBZ	M-Box RNA, ykoK riboswitch aptamer	161	28
4GXY	Adenosylcobalamin riboswitch	172	34
4P8Z	18S rRNA gene	188	29
4P9R	RNA (189-MER)	192	30
4GMA	Adenosylcobalamin riboswitch	210	36
4C4Q	Internal Ribosomal Entry Site	233	49
3DHS	RNase P RNA	268	42

Table 2: **Test set.** The *test set* contains 25 molecules.

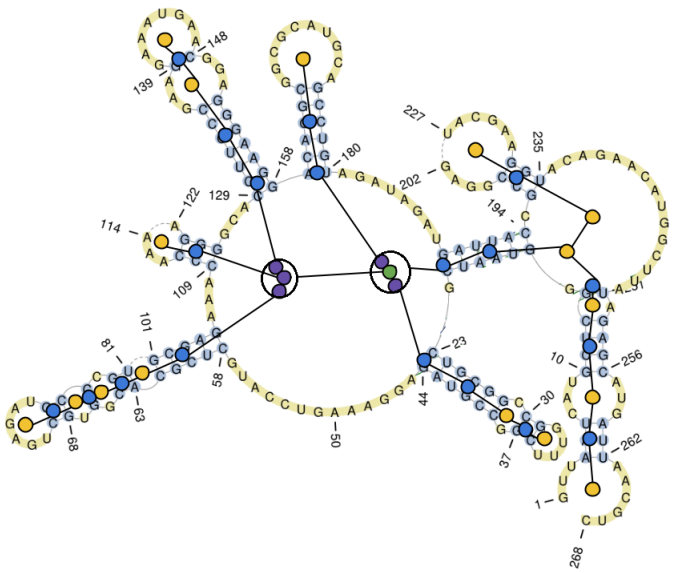

Figure 1: **Example of graph representation.** Helices are represented by one or more blue nodes. Terminal loops and 2-way junctions are represented by one yellow node. 3-way junctions are represented by two yellow nodes. 4-way junctions and more are represented by purple and green nodes.

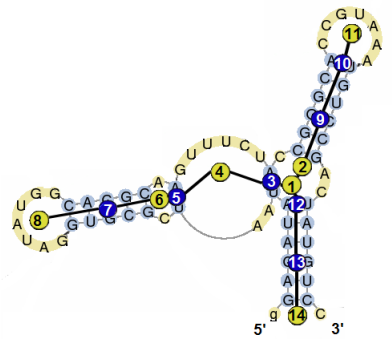

Figure 2: **Ordering of the players.** Players from the largest degree junction play first. The other players are numbered according to a depth-first search, starting from the first largest degree junction according to 5'-3' ordering.

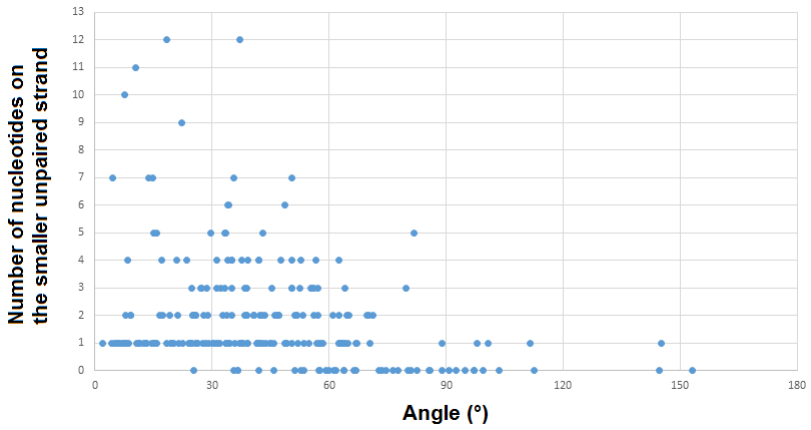

Figure 3: **Angle between two edges for 2-way junction.** The graphic represents the observed relation between the number of nucleotides on the smallest non-paired strand, and the angle between two edges of the 2-way junction nodes, for all molecules of the *reference set*. From this relation, we decided to create two set of strategies: one gives an angle between 30^o and 120^o for bulges (that is with 0 nucleotides on the non-paired strand), the other one gives an angle between 0^o and 30^o for all others cases.

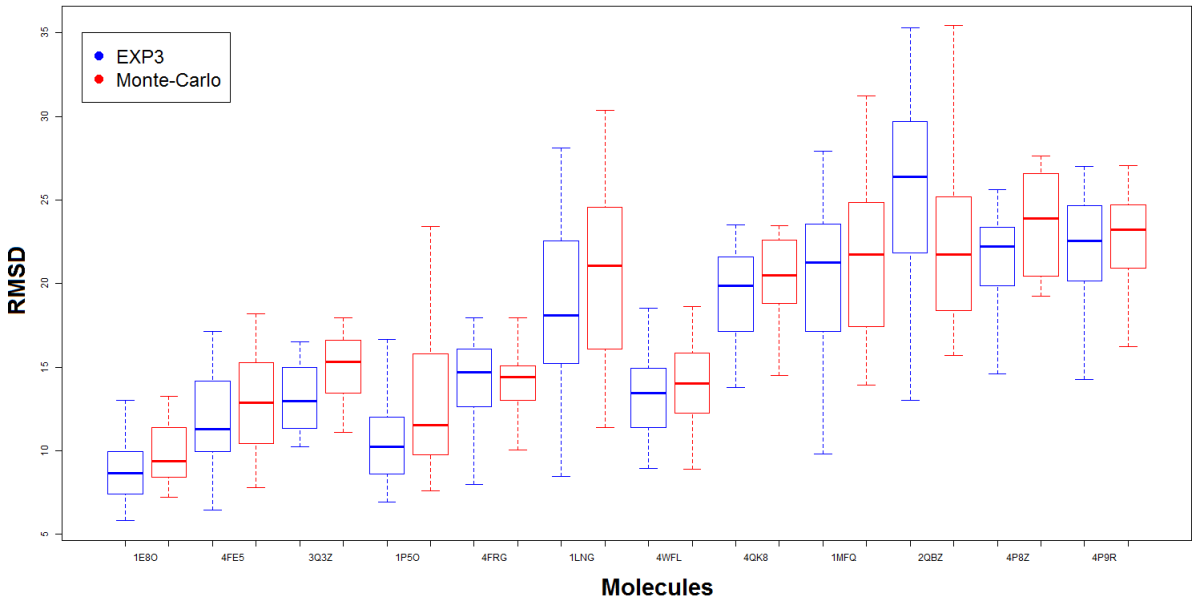

Figure 4: **Comparison between EXP3 algorithm and Monte-Carlo method.** We substituted the EXP3 algorithm of GARN2 with a Monte-Carlo method. For molecules from the *test set*, the sampling of Monte-Carlo is less efficient than EXP3 sampling: the RMSD mean is higher for most structures given by Monte-Carlo, and the minimal RMSD is always in the EXP3 sampling.

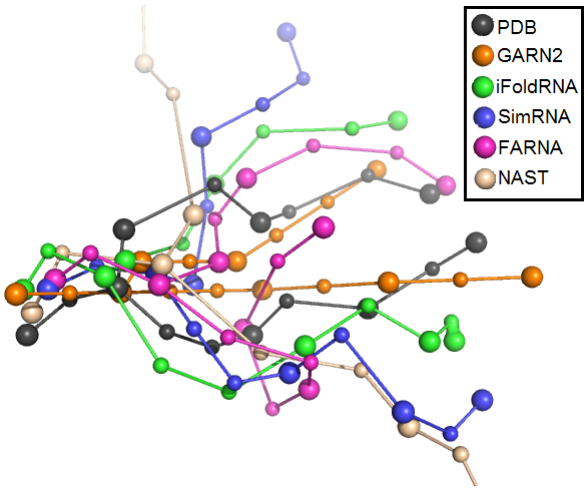

Figure 5: **Example of best results for the molecule 1LNG.** This figure shows the best solution for the fine-grained methods and GARN2. Each fine-grained structure was translated into a GARN2 graph. Despite the rigidity of the structure, GARN2 proposes a good structure compared to other approaches, as SimRNA and NAST, which propose too extended structures.

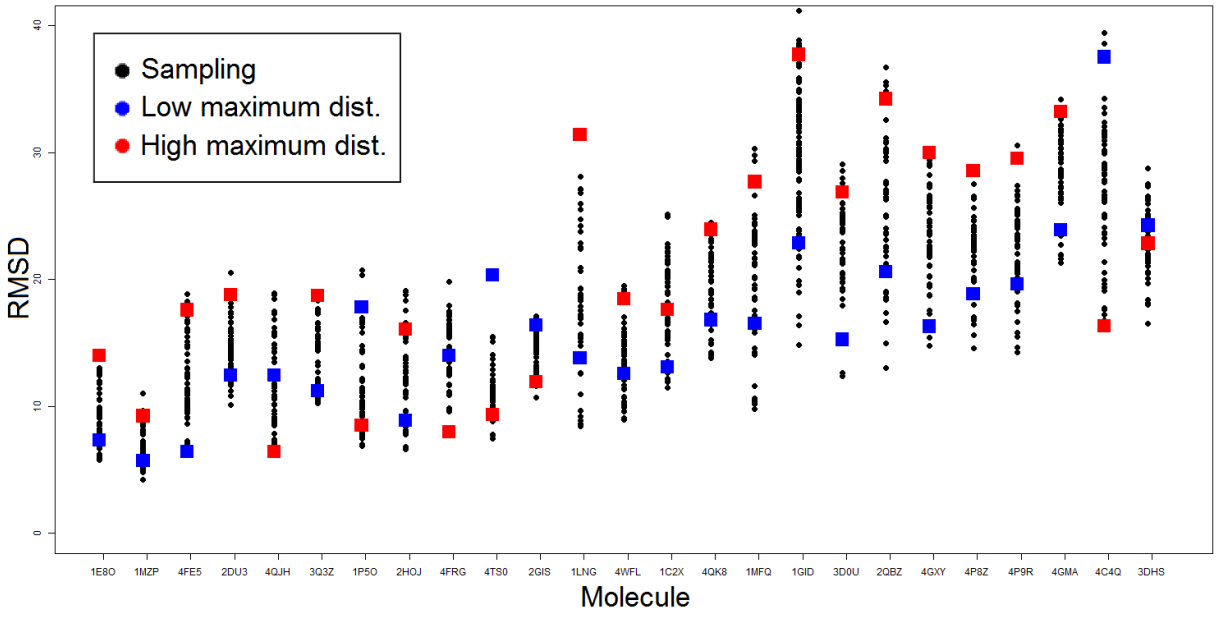

Figure 6: **Sorting of GARN sampling for the *test set*.** We have sorted the GARN sampling of the *test set*, with the maximum distance criterion. For each molecule, the two extracted structures are the one with lowest maximum distance (blue) and the one with the highest maximum distance (red) of the sampling. In most cases, one of the two structures is within the 25% best structures.

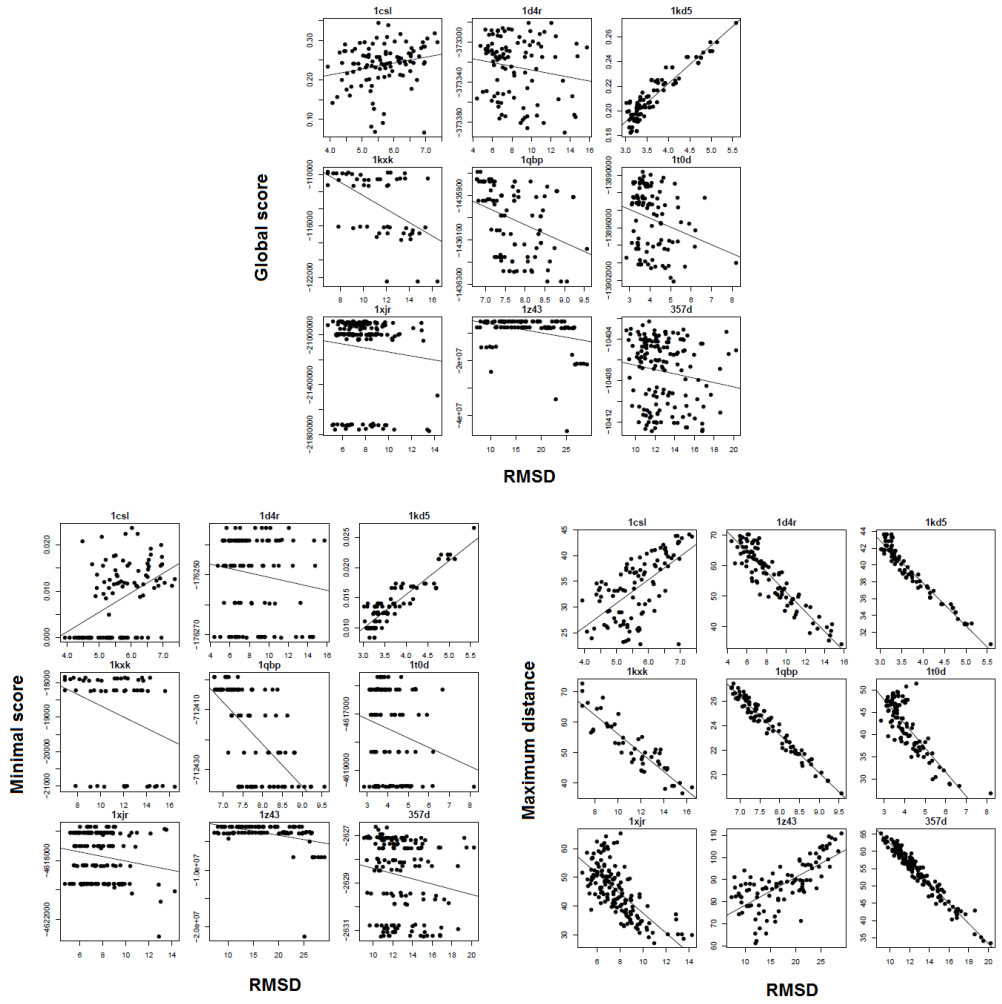

Figure 7: **Comparison of three sorting criteria on the *reference set*.** To find the best sorting criterion, we have tested three criteria on the *reference set*. For most molecules, no correlation between RMSD or score criteria (minimal score and correlation score) can be observed. On the contrary, there is a correlation between RMSD and maximum distance: we observe a linear regression whose parameter can be be either positive or negative.

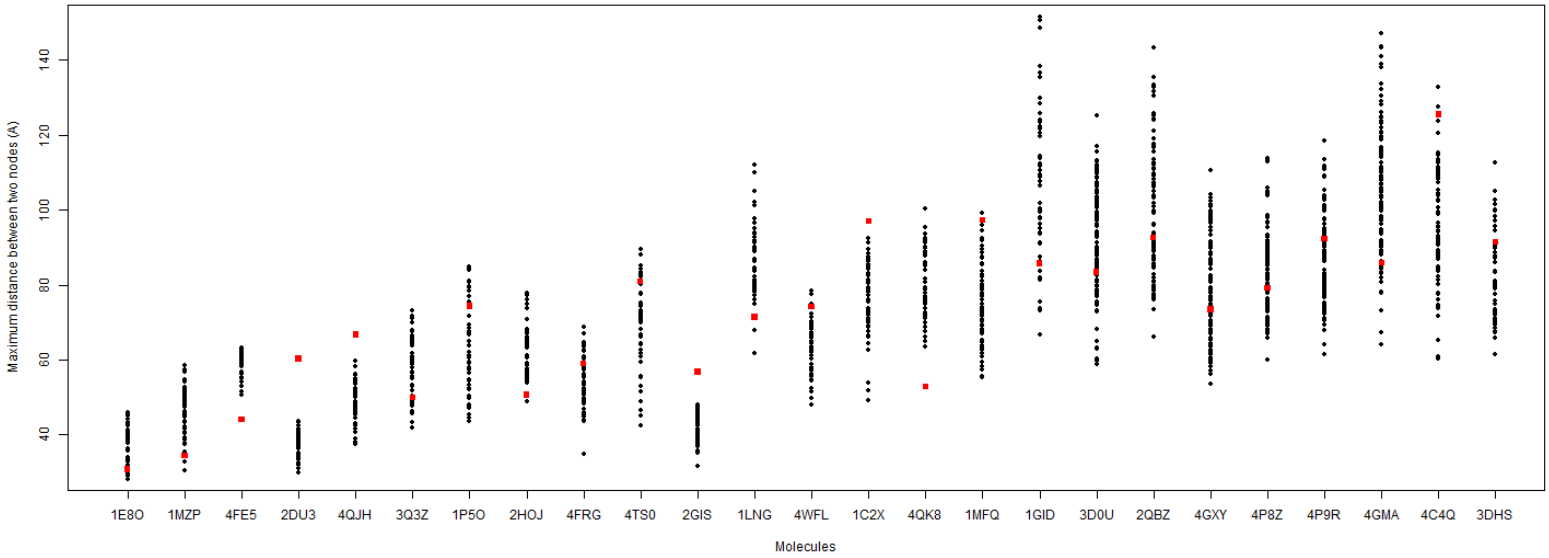

Figure 8: **Maximum distance of GARN sampling for the *test set*.** The graphic represents the GARN sampling of the *test set*, according to the maximum distance criterion (maximum distance beween two nodes in the structure). The red point corresponds to the maximum distance in the native structure.

ID PDB	Description	# Nucl.	# Players	RMSD		
				Min.	Mean	Max.
1I97	16S RRNA	1514	293	43.93	57.66	79.12
1C2W	23S Ribosomal RNA	2904	505	62.87	76.59	80.25

Table 3: **Results for very large molecules.** We tested GARN2 on two molecules with more than 1500 nucleotides. The running time has been fit to less of ∼2h per structure. For 1I97, the minimal RMSD is two times higher than the minimal RMSD of 4GMA, but the number of nucleotides is more than seven times higher. GARN2 could be seen as a first step to folding large molecules.

ID PDB	# Nucl.	# Players	RMSD			
			Min.	1^{st} quartile	Mean	Max.
1E8O	50	8	5.80	7.41	8.87	13.02
1MZP	55	8	4.20	5.95	6.98	11.03
4FE5	68	14	6.43	10.03	12.27	18.85
2DU3	71	13	10.14	13.72	14.87	20.54
4QJH	74	15	6.44	9.12	12.01	18.90
3Q3Z	77	9	10.24	11.51	13.82	18.91
1P5O	77	19	6.90	8.70	11.26	20.72
2HOJ	79	15	6.62	9.43	12.31	19.11
4FRG	84	17	7.99	12.80	14.37	19.83
4TS0	89	21	7.47	10.39	11.17	15.47
2GIS	94	21	10.72	12.62	14.07	17.13
1LNG	97	16	8.45	15.30	18.27	28.08
4WFL	107	18	8.93	11.69	13.64	19.52
1C2X	120	17	11.45	15.33	17.79	25.12
4QK8	124	20	13.79	17.14	19.41	24.50
1MFQ	127	24	9.62	17.33	20.35	30.32
1GID	158	27	16.36	26.04	30.08	37.72
3D0U	161	32	12.40	20.09	22.38	29.08
2QBZ	161	28	13.00	22.14	25.76	36.69
4GXY	172	34	14.75	21.08	23.23	29.41
4P8Z	188	29	14.60	20.01	21.65	27.55
4P9R	192	30	14.23	20.49	22.30	30.55
4GMA	210	36	21.33	27.22	28.66	34.17
4C4Q	233	49	16.32	24.68	27.07	39.43
3DHS	268	42	16.54	21.51	23.11	28.75

Table 4: **Results of GARN2 sampling on the *test set*.** For each molecule, GARN2 computes 50 possibles structures. This table shows the RMSD results of this sampling, compared to the PDB structure.

ID PDB	# Nucl.	GARN2	FARNA	NAST	ERNWIN	iFoldRNA	RNACo.	SimRNA
1E8O	50	1 min.	$\sim$ 5 min.	$\sim$ 10 min.	$\sim$ 20 min.	$\sim$ 1 h 30	2 min.	$\sim$ 5 h
4FE5	68	6 min.	$\sim$ 4 h	$\sim$ 20 min.	$\sim$ 30 min.	$\sim$ 2 h 30	3 min.	$\sim$ 9 h
4QJH	74	6 min.	$\sim$ 4 h30	$\sim$ 20 min.	–	$\sim$ 2 h	2 min.	$\sim$ 9 h
2HOJ	79	7 min.	$\sim$ 4 h	$\sim$ 25 min.	$\sim$ 40 min.	$\sim$ 2 h	3 min.	$\sim$ 8 h
4FRG	84	9 min.	$\sim$ 4 h	$\sim$ 25 min.	$\sim$ 45 min.	$\sim$ 5 h 40	3 min.	$\sim$ 3 h
4WFL	107	9 min.	$\sim$ 4 h 30	$\sim$ 25 min.	$\sim$ 1 h 20	$\sim$ 3 h	4 min.	$\sim$ 8 h
1MFQ	127	$\sim$ 30 min.	$\sim$ 5 h	$\sim$ 25 min.	$\sim$ 1 h	$\sim$ 4 h 30	5 min.	$\sim$ 15 h
4P8Z	188	$\sim$ 40 min.	$\sim$ 33 h	$\sim$ 40 min.	$\sim$ 4 h 40	–	7 min.	$\sim$ 18 h
4GMA	210	$\sim$ 1 h 30	$\sim$ 45 h	$\sim$ 45 min.	$\sim$ 3 h 30	–	8 min.	–
3DHS	268	$\sim$ 2 h 10	$\sim$ 70 h	$\sim$ 50 min.	$\sim$ 4 h 15	–	12 min.	–

Table 5: **Running times.** This table shows the running time for different methods. GARN2, FARNA, NAST and ERNWIN were tested on the same machine (HT Quad Core, 2.8 GHz Turbo, 10MB, 1066MHz, 8GB of RAM). iFoldRNA , RNAComposer and SimRNA were tested on their own servers. RNAComposer connects to databases (RNAFRABASE and RCSB PDB database) to rebuild the molecule with the knowledge they already have. We have no information about the running time of RNAJAG. The running times were performed for computing 50 structures. For RNAComposer, the calculation time is for 10 structures (because the server does not offer more structures). We do not provide tertiary interactions to NAST, which then does not operates long distance interactions, allowing it to be faster. The parameters used in the different approaches are the basic parameters of the software.

ID PDB	FARNA			NAST			iFoldRNA			RNAComposer		
	1^{st} **Q.**	**Low**	**High**	1^{er} **Q.**	**Low**	**High**	1^{er} **Q.**	**Low**	**High**	1^{er} **Q.**	**Low**	**High**
1E8O	11.23	**9.74**	17.72	19.34	**17.44**	23.49	12.95	**11.56**	21.69	0.78	**0.74**	1.82
1MZP	7.92	**7.60**	14.37	16.05	**15.55**	18.45	13.62	**7.73**	19.09	2.16	**1.88**	4.2
4FE5	12.33	**12.21**	19.39	22.92	**17.37**	26.2	15.29	**13.97**	17.11	1.84	2.71	2.4
2DU3	11.36	12.73	**7.96**	18.89	**17.79**	131.06	11.48	15.69	12.67	1.69	2.47	2.67
4QJH	11.95	15.41	15.29	23.24	**23.07**	24.77	15.77	**13.96**	20.58	10.15	11.45	**9.47**
3Q3Z	13.38	**12.97**	20.95	30.78	**27.95**	33.92	11.13	**10.26**	18.66	9.33	9.37	9.94
1P5O	11.67	22.7	**11.50**	17.11	**16.38**	18.93	18.97	20.18	**15.26**	4.99	6.82	7.86
2HOJ	12.72	**10.45**	21.69	26.88	**26.15**	29.96	15.04	**13.70**	18.11	17.13	**16.49**	19.65
4FRG	13.16	15.86	**9.54**	29.72	**29.29**	31.42	16.22	17.06	22.87	2.57	2.6	3.61
4TS0	10.44	16.48	12.52	–	–	–	–	–	–	8.16	8.72	9.79
2GIS	15.9	**11.78**	20.22	38.8	**37.84**	39.97	17.67	**15.36**	28.78	4.55	9.9	**4.14**
1LNG	20.27	**15.78**	32	45.47	**44.58**	47.21	16.99	**11.96**	30.44	6.69	**6.59**	7.24
4WFL	15.98	**15.04**	20.53	44.13	**43.28**	47.02	19.99	21.47	**15.12**	6.69	7.23	**6.49**
1C2X	–	–	–	–	–	–	15.57	19.67	18.76	6.08	8.72	6.59
4QK8	18.49	**14.08**	22.93	54.8	**54.29**	57.2	19.62	**18.95**	27.51	12.66	12.66	14.06
1MFQ	20.18	29.57	27.76	57.79	**57.65**	60.83	26.79	27.49	**24.37**	5.24	6.13	7.04
1GID	22.01	**21.70**	45.73	83.40	**80.57**	86.51	32.68	**29.86**	42.73	6.79	11.33	17.43
3D0U	–	–	–	79.46	**78.36**	82.33	38.33	**29.71**	45.48	7.46	**3.47**	13.17
2QBZ	23.63	24.06	30.09	86.38	**84.13**	87.68	26.72	**26.08**	31.49	5.45	8.45	**4.81**
4GXY	23.33	**19.60**	32.85	91.83	**89.06**	93.04	–	–	–	8.36	**4.86**	12.29
4P8Z	22.77	**20.34**.	25.54	100.11	101.05	100.99	–	–	–	29.07	**26.78**	29.07
4P9R	25.51	**20.49**	40.4	106.49	**100.30**	108.58	–	–	–	26.47	**25.49**	26.58
4GMA	25.53	**20.53**	43.89	117.75	**117.19**	120	–	–	–	20.3	**17.20**	25.32
4C4Q	24.75	30.74	47.7	114.41	**112.94**	118.27	–	–	–	9.57	**9.16**	50.01
3DHS	25	**24.66**	38.46	163.63	**163.31**	168.09	–	–	–	7.56	8.16	**7.16**

Table 6: **Results of sorting on others methods.** We applied our sorting criterion on the sampling of others fine-grained methods. We compared the RMSD of two extracted structures (lowest or highest maximum distance) with the 1^{st} quartile RMSD of the sampling. The choice between lowest or highest maximum distance depends on the width of the sampling. For example, NAST does not fold structures, so the lowest maximum distance is the best choice to find the best structure. RNAComposer proposes a narrow sampling of 10 closes structures only, so it is difficult to sort them with our geometric criterion.

ID PDB	Structure with lowest maximum distance		Structure with highest maximum distance		RMSD between both structures	Native Distance
	RMSD	Distance	RMSD	Distance		
1E8O	**7.31**	28.22	14.0	46.00	9.36	30.78
1MZP	**5.71**	30.60	9.24	58.56	10.79	34.49
4FE5	**6.43**	50.67	17.59	63.09	16.70	44.03
2DU3	**12.47**	39.45	18.81	29.78	8.26	60.24
4QJH	12.45	37.47	**6.44**	59.60	10.45	66.75
3Q3Z	**11.23**	41.73	18.73	73.04	12.23	49.92
1P5O	17.82	43.75	**8.51**	84.83	17.56	74.27
2HOJ	**8.88**	49.00	16.07	77.70	20.27	50.65
4FRG	14.0	34.81	**7.99**	68.84	12.22	58.90
4TS0	20.35	38.23	**9.33**	89.41	18.03	80.92
2GIS	16.39	31.68	**11.94**	48.11	9.22	56.77
1LNG	**13.83**	61.69	31.4	111.89	19.57	71.38
4WFL	**12.58**	48.00	18.49	78.50	12.28	74.20
1C2X	**13.12**	92.35	17.61	49.18	13.26	96.92
4QK8	**16.81**	63.50	23.94	100.41	18.29	52.78
1MFQ	**16.51**	55.35	27.69	99.07	21.49	97.20
1GID	**22.88**	66.57	37.72	151.45	34.52	85.66
3D0U	**15.26**	58.78	26.86	125.16	24.11	83.38
2QBZ	**20.58**	66.06	34.23	143.33	26.86	92.57
4GXY	**16.3**	53.68	30	110.59	24.76	73.44
4P8Z	**18.88**	59.88	28.58	113.70	28.23	79.16
4P9R	**19.65**	61.35	29.52	118.34	21.41	92.24
4GMA	**23.92**	64.14	33.22	156.20	23.52	85.90
4C4Q	37.52	60.25	**16.32**	132.64	31.95	125.50
3DHS	24.26	61.38	**22.84**	112.68	22.43	91.34

Table 7: **Results of sorting on GARN.** Each of the first two columns represents the RMSD between one of the two selected candidates and the native structure (left) and the maximum distance between two nodes in the candidate (right). The last two columns give, respectively, the RMSD between the two candidates and the maximum distance between two nodes in the native structure.

References

[Bernauer et al., 2011] Bernauer, J., Huang, X., Sim, A. Y. L., and Levitt, M. (2011). Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA*, 17(6):1066–1075.