Supplemental Material:

Quantitative Understanding of SHAPE Mechanism from RNA Structure and Dynamics Analysis

Travis HURST*, Xiaojun XU*, Peinan ZHAO, Shi-Jie CHEN‡

Department of Physics, Department of Biochemistry, and University of Missouri Informatics Institute, University of Missouri, Columbia, MO 65211

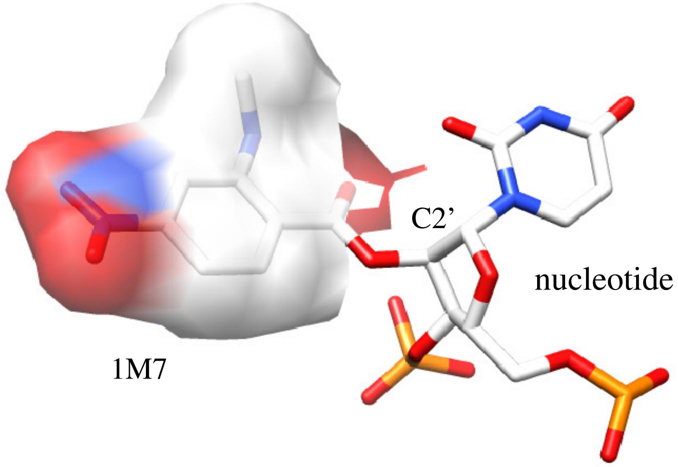

Figure S1: The 1M7-nucleotide bound conformation, built based on two similar ligands in the PDB database (PubChem CID 12076 and 66345). The effective ligand radius is measured using the Van der Waals surface shown in the figure.

Table S1: Parameter optimization using the leave-one-out approach. E_{st} and b are in kcal/mol, S_0 is in $\mathring{A}^2$, and a is dimensionless.

Value of a	Value of b	Value of E_{st}	Value of S_0	Training set	Optimal total correlation
0.03	-0.80	-0.20	16	without 3DIG	7.53
0.03	-0.45	-0.20	17	without 1Y26	7.29
0.03	-0.40	-0.25	16	without 2GDI	7.26
0.03	-0.50	-0.25	14	without 3IWN	7.48
0.03	-0.48	-0.28	14	without 1NBS	7.61
0.03	-0.60	-0.25	15	without 4KQY	7.45
0.03	-0.55	-0.25	15	without 2L8H	7.20
0.03	-0.33	-0.25	22	without 2L1V	7.46
0.03	-0.50	-0.20	15	without 3PDR	7.58
0.03	-0.45	-0.25	14	without 1VTQ	7.48
0.03	-0.48	-0.25	16	without 1EHZ	7.42
0.03	-0.43	-0.25	12	without 1AUD	7.31

Table S2: Experimental conditions of the SHAPE chemical probing for twelve RNAs.

Source name	PDB id	Reagent	Solution conditions	Ref.
Lysine riboswitch, T. maritime	3DIG	1M7	100mM HEPES, 100mM NaCl, 10mM $MgCl_2$, 310K	[2]
adenine riboswitch	1Y26	1M7	50 mM HEPES, 200mM KOAc, 3mM $MgCl_2$, 298K	[1]
TPP riboswitch, E. coli	2GDI	1M7	100mM HEPES, 100mM NaCl, 10mM $MgCl_2$, 310K	[2]
cyclic-di-GMP riboswitch, V. cholerae	3IWN	1M7	50mM HEPES, 200mM KoAc, 3mM $MgCl_2$, 310K	[2]
SAM I riboswitch, T. tengcongensis	4KQY	1M7	100mM HEPES, 100mM NaCl, 10mM $MgCl_2$, 310K	[2]
RNaseP specificity domain	1NBS	1M7	50 mM HEPES, 200 mM KOAc, 3 mM $MgCl_2$, 298K	[1]
TAR RNA from HIV-1	2L8H	1M7	50 mM HEPES, 50 mM KCl, 298K	[3]
Pre-Q1 riboswitch	2L1V	1M7	100 mM HEPES, 100 mM NaCl, 10 mM MgCl, 310K	[2]
M-Box riboswitch	3PDR	1M7	100 mM HEPES, 100 mM NaCl, 10 mM MgCl, 310K	[2]
tRNAasp	1VTQ	1M7	100 mM HEPES, 100 mM NaCl, 10 mM MgCl, 310K	[2]
tRNAphe	1EHZ	1M7	100 mM HEPES, 100 mM NaCl, 10 mM MgCl, 310K	[4]
U1A protein binding site RNA	1AUD	1M7	50 mM HEPES, 50 mM KCl, 298K	[3]

Table S3: Sequences, native 2D, and non-native 2D structures.

PDB id	Label	Structure
1AUD	Sequence	GGCAGAGUCCUUCGGGACAUUGCACCUGCC
	Native	(((((.((((....)))).......)))))
	Non-native 1	.((((.((((....)))).)))).......
	Non-native 2	......((((....))))...((....)).
	Non-native 3	(((((.((((....)).....))..)))))
	Non-native 4	(((((.((....(((....))).)))))))
	Non-native 5	((....((((....)))).)).........
1EHZ	Sequence	GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUC CACAGAAUUCGCACCA
	Native	(((((((..((((........)))).((((.........)))).....(((((.......))))))))))))....
	Non-native 1	(((((......((((((.((......)).))))))...((((..((((........)))) ..)))).)))))....
	Non-native 2	((((((((..(((....))).((((((..((((...........))))..)))))).... ....))))))))....
	Non-native 3	((((((((...((((((.((......)).)))))).((((.(((.(.....).))))))) ....))))))))....
	Non-native 4	((((((((........(((((((((((..((((...........))))..))))))..))).))))))))))....
	Non-native 5	((((((((.((((........))))....((.((((.((..((....)).)).)))).)) ....))))))))....
1NBS	Sequence	GCGAGCCUAGCGAAGUCAUAAGCUAGGGCAGUCUUUAGAGGCUGACGGCAGGAAAAAAGC CUACGUCUUCGGAUAUGGCUGAGUAUCCUUGAAAGUGCCACAGUGACGAAGUCUCACUAG AAAUGGUGAGAGUGGAACGCGGUAAACCCCUCGC
	Native	(((((((((((..........)))))).(((((......)))))..((.((((((..((((...(((....)))..))))...)))))).....(((((............((((((((. ...)))))))).........)))..)))))))))
	Non-native 1	(((((((((((..........)))))).((((((....))))))..((.((((....((((...(((....)))..)))).....))))......((((...((..((...((((((((. ...)))))))).))..))..))))..)).)))))
	Non-native 2	(((((((((((..........)))))).((((((....))))))((.((.(((....((((...(((....)))..)))).....))).........((((..........((((((((. ...))))))))))))...)).))......)))))
	Non-native 3	(((((((((((..........)))))).((((((....))))))..(((((((....((((...(((....)))..)))).....))).......))))...(((......((((((((. ...))))))))......))).........)))))
	Non-native 4	(((((....(((..(((((.........((((((....))))))..(((((((....((((...(((....)))..)))).....))).......))))...)))))....((((((((. ...))))))))......)))((....)).)))))
	Non-native 5	(((((((((((..........))))))((..(((......((((..(((((((....((((...(((....)))..)))).....))).......)))).)))).......((((((((. ...))))))))..)))..)).........)))))

Table S4: Sequences, native 2D, and non-native 2D structures.

PDB id	Label	Structure
1VTQ	Sequence	UCCGUGAUAGUUUAAUGGUCAGAAUGGGCGCUUGUCGCGUGCCAGAUCGGGGUUCAAUUCCCCGUCGCGGAGCCA
	Native	(((((((..((((........))))((((((.......))))))...(((((.......)))))))))))))....
	Non-native 1	((((((((........((((....(((((((.....)))).)))))))((((......)))).))))))))....
	Non-native 2	((((((((.......((((........((((.....))))))))....((((......)))).))))))))....
	Non-native 3	((((((((.......((((....(((((......)).)))))))....((((......)))).))))))))....
	Non-native 4	((((((((.(((...((((........((((.....))))))))))).((((......)))).))))))))....
	Non-native 5	((((((((.........)))).....(((((.......))))).(((.((((......)))).))).))))....
1Y26	Sequence	CGCUUCAUAUAAUCCUAAUGAUAUGGUUUGGGAGUUUCUACCAAGAGCCUUAAACUCUUGAUUAUGAAGUG
	Native	(((((((((...((((((......[[.))))))........((((((]].....))))))..)))))))))
	Non-native 1	(((((((((...................((((....)))).((((((.......))))))..)))))))))
	Non-native 2	(((((((((....((.........)).((((((((((((......))....)))))))))).)))))))))
	Non-native 3	(((((((((..(((.....)))..((((((((.((((.......))))))))))))......)))))))))
	Non-native 4	(((((((((..(((.........((((..(((....))))))).(((.......)))..))))))))))))
	Non-native 5	(((((((((..(((..........(((((((.........))))..)))..........))))))))))))
2GDI	Sequence	GACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUGAUCUGGAUAAUGCCAGCGUAGGGAAGUUC
	Native	(((((((((.(((.(((.....))))))........)))))....((((...((((......))))...)))).)))).
	Non-native 1	((..((((..(((.(((.....)))))).........))))..)).((((..((((......))))..)))))......
	Non-native 2	..(((((.....(((((.....))))))))))...........((.((((..((((......))))..))))))).....
	Non-native 3	((..((((.....(((..((......)).))).....))))..)).((((..((((......))))..)))))......
	Non-native 4	((((..(((((.(((((.....)))))........)))))......((((..((((......))))..))))..)))).
	Non-native 5	..(((((.....(((((.....))))))))))).....(((..(((....)))((((......))))....)))......

Table S5: Sequences, native 2D, and non-native 2D structures.

PDB id	Label	Structure
2L1V	Sequence	`AGAGGUUCUAGUUAUACCCUCUAUAAAAAACUAA`
	Native	`(((((...[[[......)))))........]]].`
	Non-native 1	`(((((............)))))............`
	Non-native 2	`(((((...((....)).)))))............`
	Non-native 3	`...((((....(((((.....)))))..))))..`
	Non-native 4	`(((((............))))).((......)).`
	Non-native 5	`........(((((...............))))).`
3DIG	Sequence	`GGCCGACGGAGGCGCGCCCGAGAUGAGUAGGCUGUCCCAUCAGGGGAGGAAUCGGGGACGGCUGAAAGGCGAGGGCGCCGAAGGGUGCAGAGUUCCUCCCGCUCUGCAUGCCUGGGGGUAUGGGGAAUACCCAUACCACUGUCACGGAGGUCUCUCCGUGGAGAGCCGUCGGUC`
	Native	`(((((((((....(((((((((.....(((((.(((((.....[[[[[[.....))))).)))))....)))))))))...((((((((((((.]]]]]])))))))))).)))(((((((((((((...))))))))))))(((((((((....)))))))))...)))))))))`
	Non-native 1	`(((((((((.(((((.((((.........(((((((((.....((......)).)))))))))......)).)))))))..((((((((((((.......))))))))).)))...((((((((.....)))))))).((.((((((((....))))))))))..)))))))))`
	Non-native 2	`(((((((((.(((..((((..........(((((((((.....((......)).))))))))....((((....))))...))))(((((((.......)))))))..)))..((((((((((.....)))))))).)).((((((((....))))))))....)))))))))`
	Non-native 3	`(((((((((.((((.((..(((.......(((((((((.....((......)).))))))))....((.(((((((((....))))))......)))))).))).)).))))..((((((((((.....)))))))).)).((((((((....))))))))....)))))))))`
	Non-native 4	`(((((((((.((((.((..(((..(((..(((((((((.....((......)).))))))))....(((....(((((....)))))...))).)))...))).)).))))..((((((((((.....)))))))).)).((((((((....))))))))....)))))))))`
	Non-native 5	`((((((((((((.((((((..........(((((((((.....((......)).))))))))....((((....))))...))))))......))))).(((((.........((((((((((.....)))))))).))..(((((((....))))))).))))).)))))))`
3IWN	Sequence	`CACGCACAGGGCAAACCAUUCGAAAGAGUGGGACGCAAAGCCUCCGGCCUAAACGGCAUUGCACUCCGCCGUAGGUAGCGGGGUUACCGAUGG`
	Native	`((((......(((((((((((....)))))))))))...(((.((((((((..((((..........)))))))))..))))))...)).)).`
	Non-native 1	`....((..((((...((((((....))))))...))..((((.(((((((..(((((..........)))))))))..))))))).))..)).`
	Non-native 2	`((((......((...((((((....))))))...))..((((.(((((((..(((((..........)))))))))..)))))))..)).)).`
	Non-native 3	`...((.....))...((((((....))))))..((...((((.(((((((..(((((..........)))))))))..)))))))..))....`
	Non-native 4	`.........(((...((((((....))))))...((((.(((............))).)))).....)))...((((((...)))))).....`
	Non-native 5	`....((..((.....((((((....))))))...((((.(((............))).)))).((((((........))))))...))..)).`

Table S6: Sequences, native 2D, and non-native 2D structures.

PDB id	Label	Structure
2L8H	Sequence	GGCAGAUCUGAGCCUGGGAGCUCUCUGCC
	Native	((((((...((((......))))))))))
	Non-native 1	((((((...(((.((...)))))))))))
	Non-native 2	(((((((((.......)))....))))))
	Non-native 3	(((((((((......))).....))))))
	Non-native 4	((((((((((....)))).....))))))
	Non-native 5	((((((.((.........))...))))))
3PDR	Sequence	CUUCGUUAGGUGAGGCUCCUGUAUGGAGAUACGCUGCUGCCCAAAAAUGUCCAAAGACGCCAAUGGGUCAACAGAAAUCAUCGACAUAAGGUGAUUUUUAAUGCAGCUGGAUGCUUGUCCUAUGCCAUACAGUGCUAAAGCUCUACGAUUGAAG
	Native	((((((..(((..(((..((((((((......(((((.(((((....((((....)))).....))))).....(((((((((.......)))))))))....))))).((........))....)))))))).)))...)))..))....))))
	Non-native 1	..((((..(((..(((..((((((((......(((((.(((((....((((....)))).....)))))....((((((((((.......))))))))))...))))).((((....))))....)))))))).)))...)))..))))......
	Non-native 2	..((((..(((..(((..((((((((......((((((..................(((.((...)))))...((((((((((.......))))))))))...))))).((((....))))....)))))))).)))...)))..))))......
	Non-native 3	..((((..(((..(((..((((((((......(((((..........((((((..........)))).))..((((((((((.......))))))))))...))))).((((....))))....)))))))).)))...)))..))))......
	Non-native 4	(((((.....)))))...((((((((......(((((.(((((....((((....)))).....)))))....((((((((((.......))))))))))...))))).((((....))))....)))))))).((....)).............
	Non-native 5	((((....(((....((((.....))))....(((((.(((((....((((....)))).....)))))....((((((((((.......))))))))))...))))).((((....))))...)))....((.((....)).))......))))
4KQY	Sequence	UUCUUAUCAAGAGAAGCAGAGGGACUGGCCCGACGAAGCUUCAGCAACCGGUGUAAUGGCGAAAGCCAUGACCAAGGUGCUAAAUCCAGCAAGCUCGAACAGCUUGGAAGAUAAGAA
	Native	((((((((......(((...(((.[[[[)))......)))..(((..(((((...(((((....))))).)))..)).)))....]]]](((((.......)))))...))))))))
	Non-native 1	.(((((((....(((((...(((.....)))......))))).(((.((...((.(((((....))))).))...))))).........((((((.....))))))...))))))).
	Non-native 2	.(((((((....(((((...(((.....)))......))))).(((.((...((.(((((....))))).))...)))))....((((...((((.....))))))))).))))))).
	Non-native 3	.(((((((............(((..((((((......((....))....(((...(((((....))))).)))..)).))))..)))..((((((.....))))))...))))))).
	Non-native 4	.(((((((.....((((...(((.....)))..(((.((((.((((.((...((.(((((....))))).))...)))))).........)))))))....))))....))))))).
	Non-native 5	.(((((((.............(((..(((........)))..((((.((...((.(((((....))))).))...))))))...)))..((((((.....))))))...))))))).

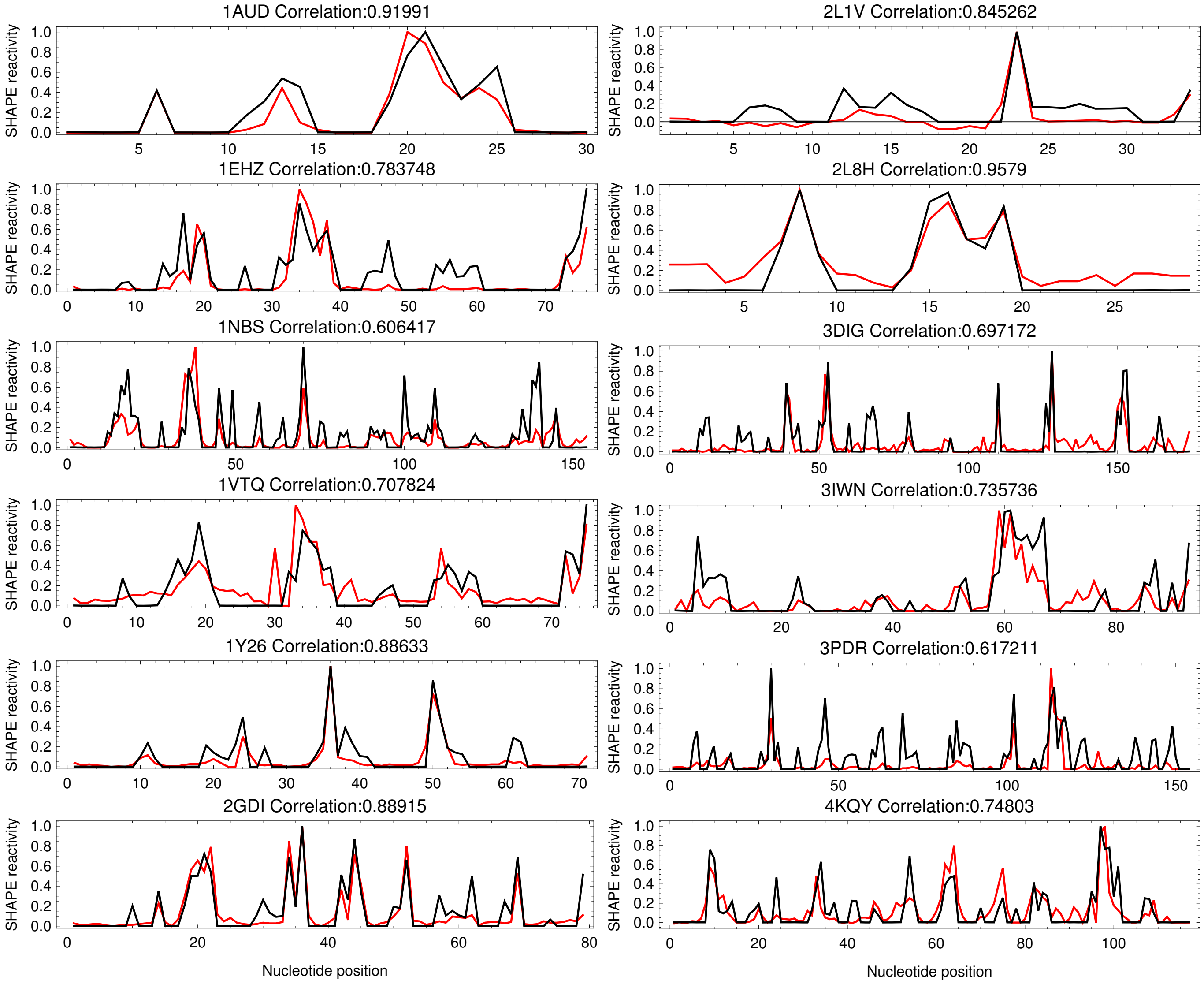

Figure S2: Comparison between the predicted (black) and the experimental (red) SHAPE profiles.

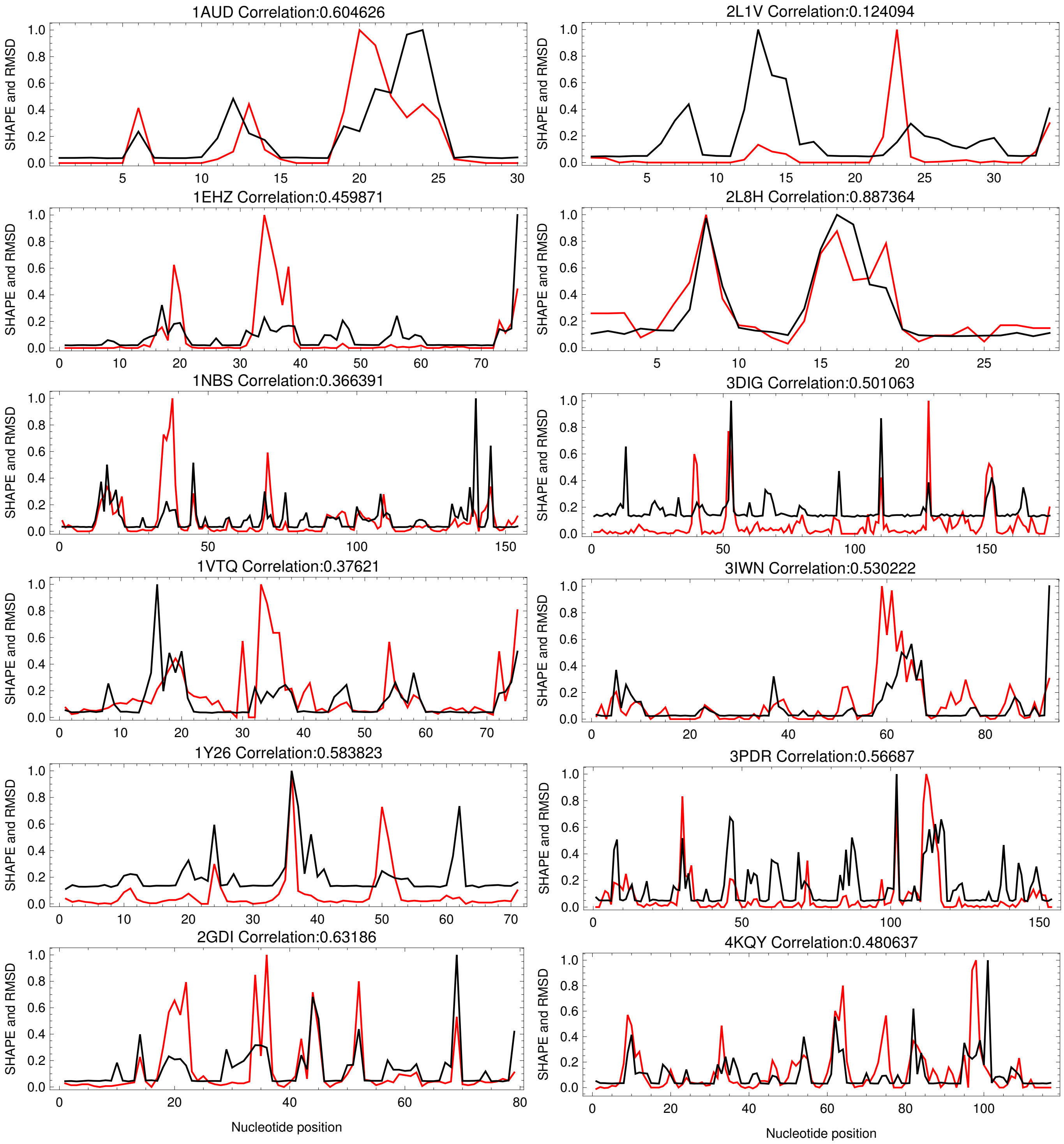

Figure S3: Comparison between the nucleotide RMSD (black) and the experimental SHAPE (red) profiles.

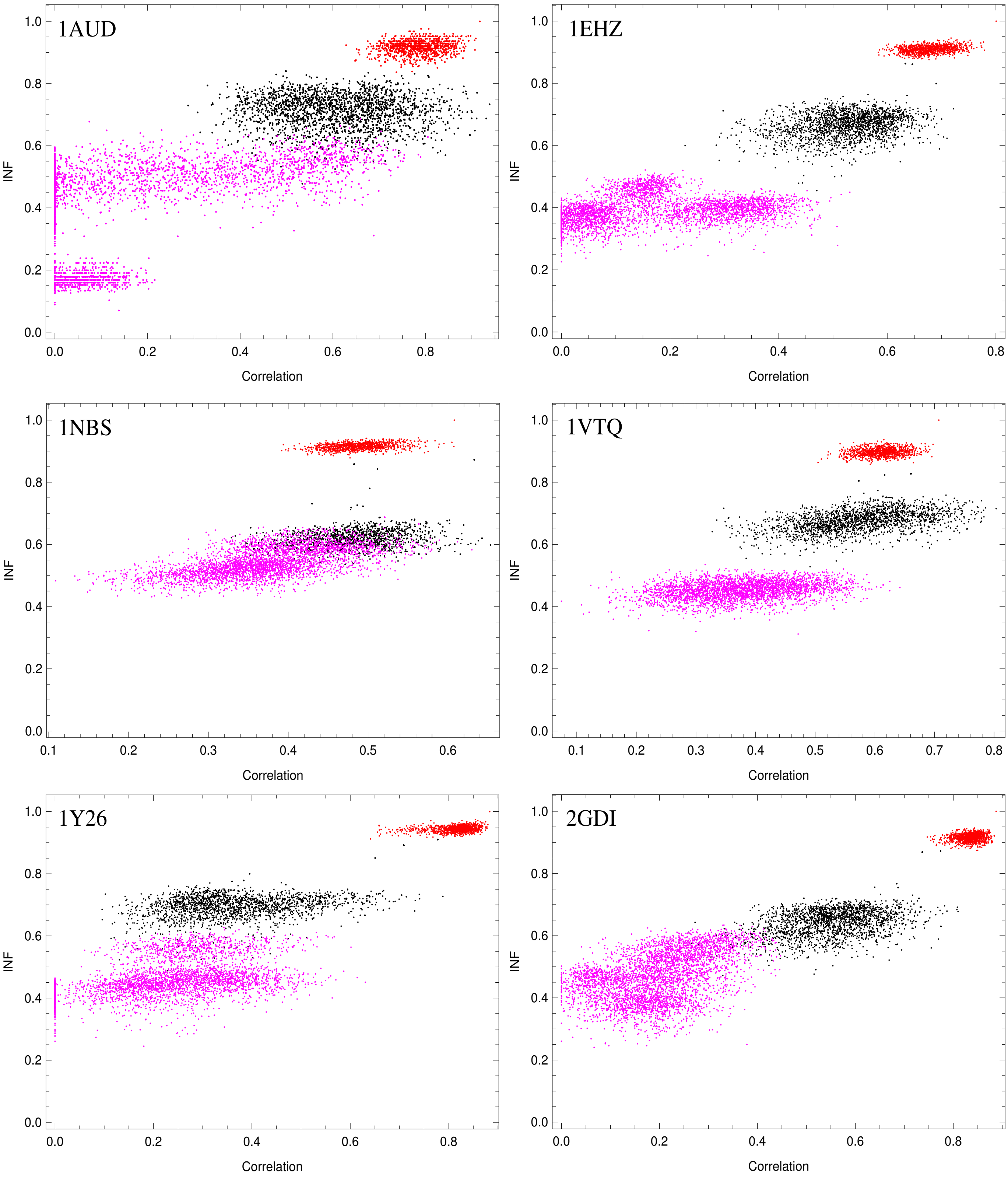

Figure S4: Comparison of INF and correlation for the first six cases. Near-native structures from all-atom MD sampling are red; native 2D, non-native 3D structures from CG sampling are black; non-native 2D, non-native 3D structures from CG sampling are magenta.

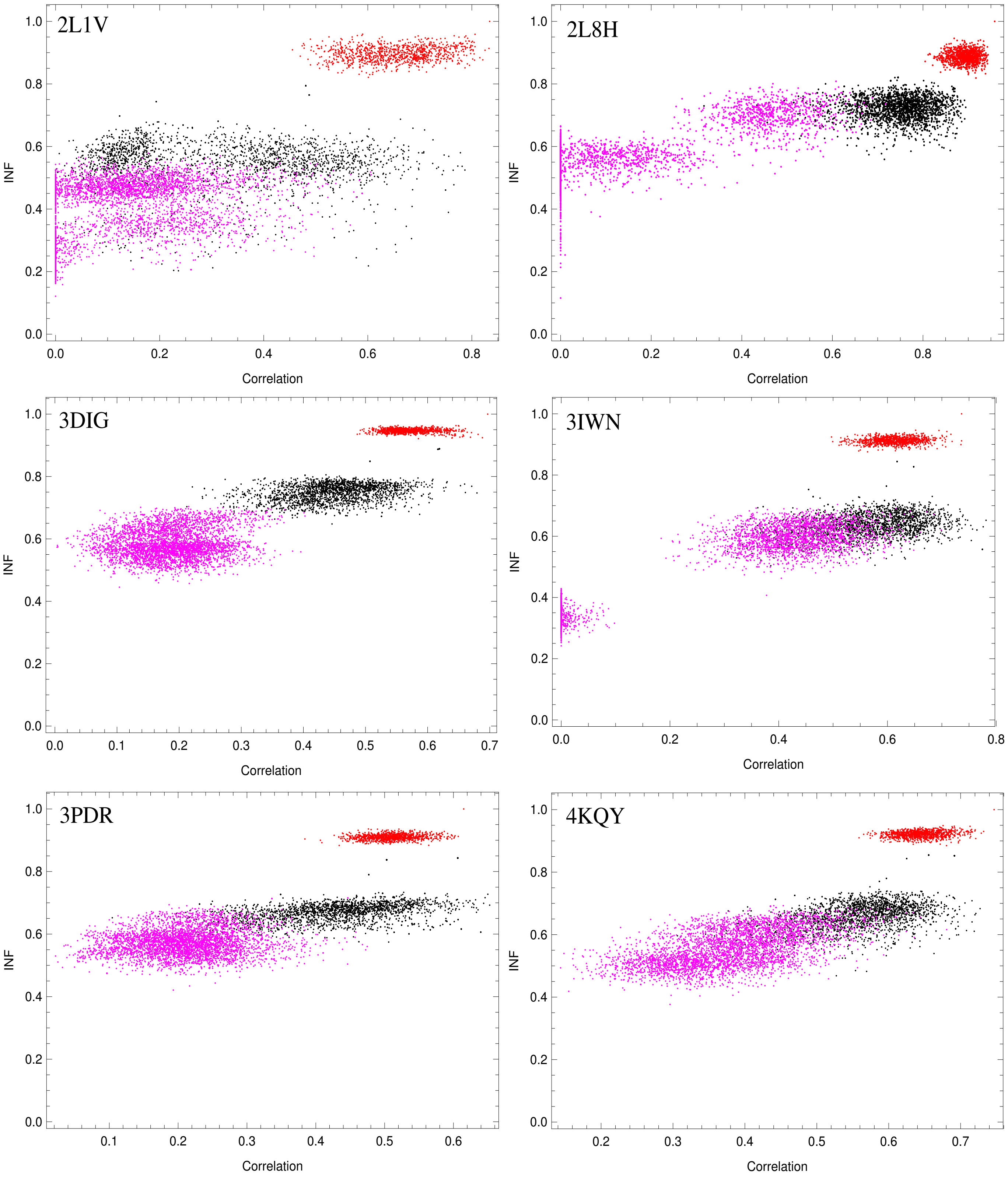

Figure S5: Comparison of INF and correlation for the last six cases. Near-native structures from all-atom MD sampling are red; native 2D, non-native 3D structures from CG sampling are black; non-native 2D, non-native 3D structures from CG sampling are magenta.

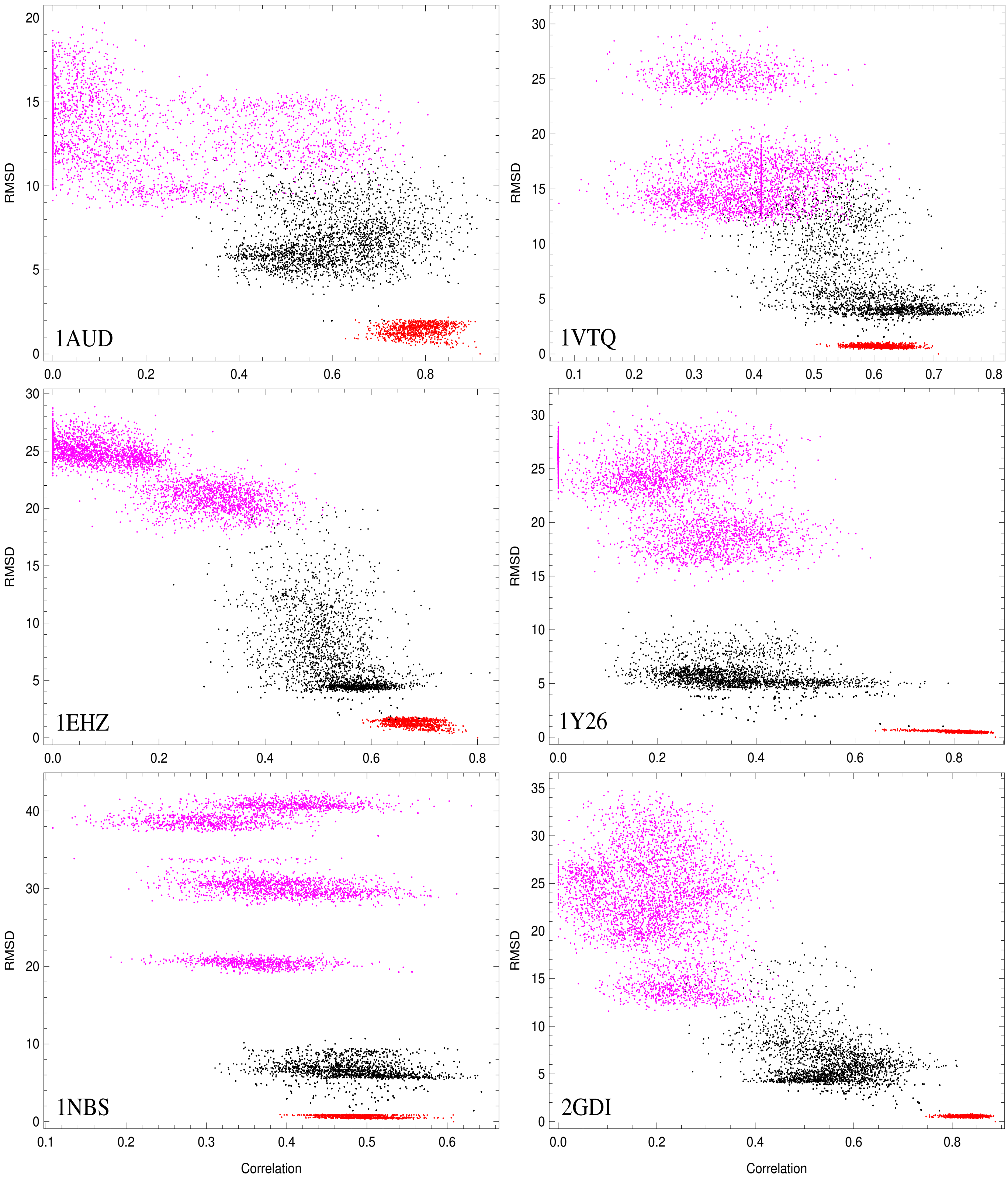

Figure S6: Comparison of RMSD and correlation for the first six cases. Near-native structures from all-atom MD sampling are red; native 2D, non-native 3D structures from CG sampling are black; non-native 2D, non-native 3D structures from CG sampling are magenta.

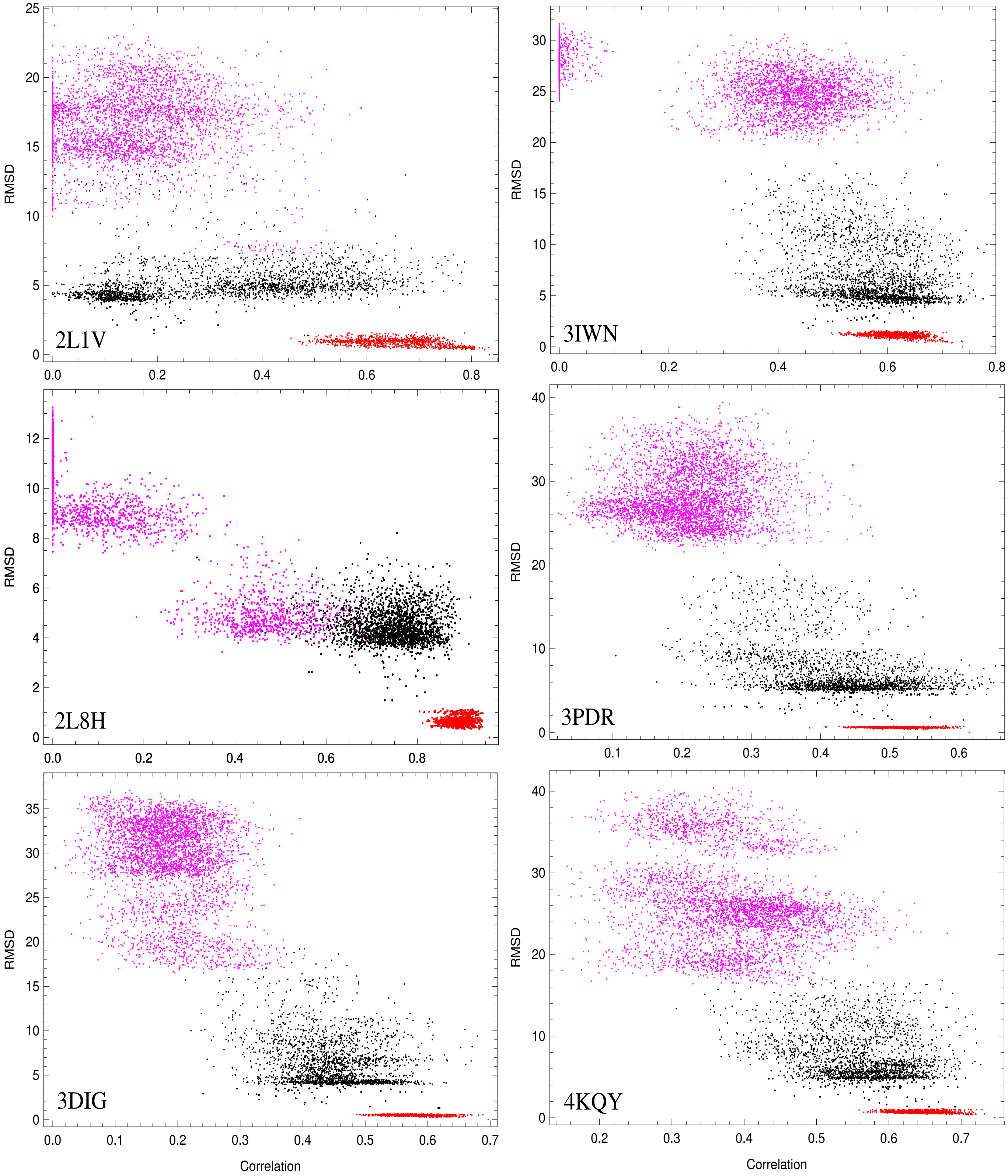

Figure S7: Comparison of RMSD and correlation for the last six cases. Near-native structures from all-atom MD sampling are red; native 2D, non-native 3D structures from CG sampling are black; non-native 2D, non-native 3D structures from CG sampling are magenta.

References

[1] Leonard, C. W.; Hajdin, C. E.; Karabiber, F.; Mathews, D. H.; Favorov, O. V.; Dokholyan, N. V.; Weeks, K. M. Principles for understanding the accuracy of SHAPE-directed RNA structure modeling. *Biochemistry*, **2013**, *52*, 588-595.

[2] Hajdin, C. E.; Bellaousov, S.; Huggins, W.; Leonard, C. W.; Mathews, D. H.; Weeks, K. M. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U. S. A.*, **2013**, *110*, 5498-5503.

[3] Gherghe, C. M.; Shajani, Z.; Wilkinson, K. A.; Varani, G.; Weeks, K. M. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S^2) in RNA. *J. Am. Chem. Soc.*, **2008**, *130*, 12244-12245.

[4] Rice, G. M.; Leonard, C. W.; Weeks, K. M. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA*, **2014**, *20*, 846-854.