Supporting Information

Structure and Thermodynamics of N6-Methyladenosine in RNA: a Spring-loaded Base Modification

Caroline Roost[1], Stephen R. Lynch[1], Pedro J. Batista[2], Kun Qu[2], Howard Y. Chang[2*], and Eric T. Kool[1,*]

[1]*Department of Chemistry, Stanford University, Stanford, California 94305*
[2]*Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford, CA 94305.*
*Authors to whom correspondence should be addressed: kool@stanford.edu; howchang@stanford.edu

Contents:

Experimental Methods

Oligonucleotide Synthesis. RNA oligonucleotides were synthesized by automated methods on an Applied Biosystems 394 DNA/RNA synthesizer (1 μmol scale) in "DMT On" mode, using standard β-cyanoethyl phosphoramidite chemistry and 2'-*O*-TBDMS-protected ribonucleosides. The m^6A phosphoramidite was synthesized according to literature procedures.[1] Phosphoramidites of A, C, G, and U and synthesizer reagents were purchased from Glen Research. Cleavage from CPG solid support and deprotection of the nucleobase protecting groups were achieved with AMA solution (50:50, v/v mixture of 28% aq. ammonium hydroxide and 40% aq. methylamine) for 10 min at 65 °C. Deprotection of the 2'-*O*-TBDMS group and initial purification were carried out with Glen-Pak RNA purification cartridges according to the manufacturer's procedure. RNA oligonucleotides were further purified by 20 % denaturing polyacrylamide gel electrophoresis and isolated by excision and extraction from the gel, followed by dialysis against deionized water. Purity and identity were confirmed with MALDI-TOF mass spectrometry (Stanford Peptide and Nucleic Acid Facility). Concentrations were determined by UV absorbance at 260 nm. Molar extinction coefficients were calculated by the nearest neighbor method. The extinction coefficient of m6A was assumed to be the same as adenosine.

Thermal Denaturation Studies. Solution for the thermal melting experiments contained a 1:1 ratio of two complementary strands in 1-5 μM total RNA concentration (C_t). Self-complementary cases were measured in 2-14 μM concentrations. The oligoribonucleotides were measured in 10 mM sodium phosphate buffer, pH 7.0 with 1 M NaCl. Prior to the measurements, solutions were heated to 90 °C and then slowly cooled to 10 °C to ensure a complete annealing of the strands. Experiments were carried out in Teflon-stoppered 1 cm pathlength quartz cells under nitrogen atmosphere on a Varian Cary 100 UV-Vis spectrophotometer equipped with a thermoprogrammer. Absorbance at 260 nm was monitored while temperature was raised at a rate of 1 °C/min from 10 °C to 90 °C.

Computer fit of the first derivative of absorbance with respect to 1/T using Meltwin 3.0 provided the melting temperatures (T_m). Free energy values were determined by two methods: (1) nonlinear least squares fitting of the denaturation data, using a two-state model with linear sloping baselines and (2) van't Hoff plots by plotting $1/T_m$ vs. $\ln(C_t/4)$ by measuring T_m as a function of concentration. Close agreement was observed between the results of the two methods, indicating that the two-state approximation is reasonable for these specific sequences.

The errors in Tables 1 and 2 were calculated as follows: The error of ΔTm is the standard deviation of 3 individual measurements. The error of ΔG_{fit} is the standard deviation from 3 individual measurements. To calculate the error of ΔG_{vHoff}, the two-sigma variance of slope and intercept from the van't Hoff linear regression was determined (using Origin), which led to the error of enthalpy and entropy, respectively. Consecutively, the error of ΔG_{vHoff} was determined through error propagation. The error of ΔG_{avg} is the standard deviation of ΔG_{fit} and ΔG_{vHoff}.

NMR methods. NMR experiments were acquired on a Varian Inova 600 MHz spectrometer with a z-gradient triple resonance HCN probe, or a Varian Inova 500 MHz spectrometer with a z-gradient H{X} probe. NMR data were processed using VNMR (Agilent) software and

displayed using Sparky 3.1 (Goddard et al.).[2] ^{1}H and ^{13}C chemical shifts were referenced indirectly to tetramethylsilane (TMS), ^{31}P shifts were referenced indirectly referenced to phosphoric acid.

The exchangeable ^{1}H resonances of the 2 oligonucleotide RNA duplexes (MA and DA) were assigned using through-space homonuclear SS-NOESY experiments at 15°C with different mixing times (100 ms or 250 ms) on a sample in 90%H_20/10%D_2O.[3] The non-exchangeable ^{1}H resonances of the RNA were assigned with a combination of homonuclear NOESY experiments at varied mixing times (75 ms, 150 ms and 300 ms), DQF-COSY, homonuclear TOCSY with 80 ms mixing time, $^{1}H/^{13}C$ HSQC, and $^{1}H/^{31}P$ heteronuclear COSY acquired at 25 °C on a sample dissolved in 99.996% D_2O. All experiments were acquired on a 600 MHz Inova instrument except the $^{1}H/^{31}P$ heteronuclear COSY, which was acquired on a 500 MHz NMR instrument.

SS-NOESY spectra were acquired with 64 scans of 4096 points in t2 by 200 points in t1, with a recycle delay of 2 seconds and a spectral width of 20 ppm in both dimensions. The ^{1}H spectral width for all data acquired in D_2O was 9 ppm. NOESY spectra were acquired with 64 scans of 4096 points in t2 by 170 points in t1 with a recycle delay of 2.28 seconds for both samples. The $^{1}H/^{13}C$ HSQC was acquired centered on either the aromatic region at 140 ppm or the ribose region at 80 ppm with a ^{13}C spectral width of 40 ppm, with 64 scans of 2048 points in t2 by 80 points in t1 with a recycle delay 5.2 seconds for the aromatic or 1.5 seconds for the ribose experiment. TOCSY spectra were acquired with 16 scans of 2404 points in t2 and 256 points in t1 with a recycle delay of 1.3 seconds. DQFCOSY spectra were acquired with 16 scans of 4096 points in t2 and 256 points in t1 with a recycle delay of 1.8 seconds. $^{1}H/^{31}P$ heteronuclear COSY spectra were acquired with 272 scans of 4096 points in t2 and 80 points in t1 with a ^{31}P spectral width of 3.75 ppm, centered at 0.4 ppm with a recycle delay of 0.8 seconds for the sample with N-methyladenosine, and 144 scans of 4096 points in t2 and 128 points in t1 with a ^{31}P spectral width of 4.25 ppm, centered at 1.3 ppm for the sample with adenosine.

Structure calculations. Structures were calculated on the RNA oligonucleotides using restrained molecular dynamics followed by energy minimization using the program XPLOR-NIH version 2.35[4,5] on an iMac with Mac OS 10.9, with a force-field consisting of bond lengths, bond angles, improper angles, repulsive van der Waals potentials, and experimental distance and dihedral constraints in the absence of electrostatics. Random starting structures were created by randomizing torsion angles for the calculation. The initial stage was a modified simulated annealing protocol[6] that included experimental distance constraints but did not include dihedral constraints; structures that converged to low total energy were subjected to a refinement protocol that added in the experimental dihedral constraints.

NOE force constants were set to square wells at 50 kcal mol^{-1} $Å^2$; torsion angle force constants were varied from 10 to 50 kcal mol^{-1} rad^2. Bond length force constants were set to 1000 kcal mol^{-1}; bond angle force constants and improper force constants were set to 500 kcal mol^{-1} rad^2. The high temperature annealing was done with 15,000 cycles molecular dynamics at 1000 K with low values of interatomic repulsion scaled at 0.01, followed by 1000 steps at 0.001 ps per step of dynamics, 1.0 ps total, while increasing the van der Waals force constant to 0.1, and slow cooling with 500 steps of dynamics with 0.001 ps per step while cooling to 300 K, increasing the dihedral force constant from 10 to 50 kcal mol^{-1} rad^2. An initial energy minimization with 1000 steps with both the NOE and dihedral force constants at 50 kcal mol^{-1} rad^2 was accomplished. Of 30 initial structures, the 25 low energy structures were subjected

to a second stage of 1000 cycles of molecular dynamics at 1000 K with a time step of 0.0005 ps while increasing the torsion angle force constant from 10 to 50 kcal mol^{-1} rad^2, and 250 steps of molecular dynamics with 0.0005 ps per step while cooling to 300 K. This refinement step was completed with 1000 steps of energy minimization at 300 K. The final step on the structure calculation was a final 2000 steps of molecular dynamics at 0.001 ps per step at 300 K while increasing the force constant on the van der Waals potentials to 100, and a final 2000 steps of energy minimization that included all of the above constraints in addition to attractive Lennard–Jones potentials without electrostatic potentials. The final structures were displayed with the program PyMOL version 1.7 (The PyMOL Molecular Graphics System, Version 1.7.0.3 Schrödinger, LLC).

Distance restraints were assigned based upon NOE volumes of crosspeaks in NOESY experiments at short mixing time (75 ms), and long mixing time (150 ms). NOEs with volumes comparable to those of the H1′-H2′ and pyrimidine H5-H6 NOEs at short mixing time were assigned as strong and given a distance range of 1.8-3.0 Å. NOEs that were weaker than the strong NOEs at short mixing time but were intense at 150 ms mixing time were assigned as medium and given a distance range of 2.0-4.0 Å. The remainder of NOEs observed at 150 ms were assigned as weak and given a distance range of 3.0-5.0 Å. NOEs that were clearly identifiable but whose intensity was hard to define due to overlap were given wider ranges, reflective of what mixing time they were first observed. Water-exchangeable NOEs were assigned wide ranges based upon their intensity in 100 ms and 250 ms mixing time SS-NOESY in H_2O. Hydrogen bonds for the Watson-Crick base pairs were assigned on the basis of intense NOEs between the guanosine imino proton and the cytidine amino proton or the uridine imino proton and the adenosine H2, as well as the downfield shift of the imino proton resonance. All base pairs were modelled as Watson-Crick base pairs with distance constraints between the bases for the Hydrogen bonds.

Dihedral constraints were determined from homonuclear and heteronuclear correlation experiments, except for the χ angle, which was assigned on the basis of the presence or absence of a strong H1′-H8/H6 NOE. All nucleotides were assigned as *anti*. *Anti* was assigned a range of –40°-180° (O4′-C1′-N9/N1-C4/C2).[7] The ribose sugar pucker was assigned on the basis of the size of the H1′-H2′ coupling constant determined in a DQF-COSY. A coupling constant <2 Hz was assigned as C3′-*endo*. All nucleotides were assigned as C3′-*endo*. The values for each were from Saenger.[7]

The backbone torsion angles were essentially determined when possible as described by Marino.[8] Values for these angles were from Saenger[7] with ranges of ±20°. The backbone torsion angle ε was estimated from $^3J_{H3'\text{-}P}$ in a 2D ^{31}P-^{1}H heteronuclear COSY[9] and given a range of –155(±20)°. The torsion angle β was estimated from $^3J_{H5'\text{-}P}$ and $^3J_{H5''\text{-}P}$, from the same experiment if observed and given a range of 180(±20)°. The torsion angle γ was estimated from $^3J_{H4'\text{-}H5'}$ from a DQF-COSY for resolved nucleotides, and given a range of 55(±20)°. The final energy minimization step included these A-form torsion angle restraints ε, β, and γ for all nucleotides for both structures; α and ζ were not restrained for either structure.

BMRB Accession Code: 25291
RCSB Number: RCSB104111
PDB ID: 2MVY

BMRB Accession Code: 25220
RCSB Number: RCSB104105
PDB ID: 2MVS

RNA isolation and m6A enrichment. GM12878 cells were cultured as described in ref. 10. Cells were collected at a confluence of $6x10^5$ cells per milliliter and RNA extracted with TRIzol, according to the manufacturer's protocol. m6A immunoprecipitation and library construction was done as described by Batista et al.[11] Briefly, the RNA was re-suspended in ultrapure H_2O, treated with DNAse I (Ambion) for 30 min at 37 ºC and subjected to RNA clean up reaction with RNeasy Midi Kit (Qiagen), according to the manufacturer's protocol. RNA was eluted in ultrapure H_2O. PolyA RNA selection was performed using MicroPoly(A) Purist (Life Technologies) according to the manufacturer's protocol. For all RNA samples, the concentration, purity and integrity of the RNA were verified using a NanoDrop and Bioanalyzer. The PolyA+ RNA was fragmented to ~100 nucleotide fragments by incubation with Zinc Chloride buffer (10 mM $ZnCl_2$, 10 mM Tris-HCl, pH 7.0) for 2 minutes. The reaction was stopped with 0.2M EDTA, and the RNA precipitated with standard ethanol precipitation. 15 μg of anti-m6A polyclonal antibody (Synaptic Systems) were pretreated with agarose beads coated with ssDNA to reduce background.[12] Antibody was conjugated to Dynabeads Protein G (Life Technologies; 10003D) overnight at 4 °C. 200 μg of fragmented RNA was incubated with the antibody in 1x DamIP buffer (10 mM sodium phosphate buffer, pH 7.0, 0.3 M NaCl, 0.05% (w/v) Triton X-100) supplemented with 1% SuperRNAse Inhibitor (Ambion), for 3 hours at 4 °C. After incubation, the antibody was washed 5 times with DamIP buffer and the RNA eluted with 0.5 $mg \bullet mL^{-1}$ N6-methyladenosine (Sigma-Aldrich) in DamIP buffer.[12] One volume of ethanol was added to the eluted RNA, and the RNA recovered an RNeasy mini column. The imunoprecipitated RNA, and an equivalent amount of input RNA were used for library generation with the dUTP protocol, as described[13] except libraries were size selected by gel purification after ligation and after PCR amplification. Libraries were sequenced using an Illumina HiSeq at the Stanford Center for Genomics and Personalized Medicine. For all libraries, single-end RNA-Seq reads were mapped to the human genome (hg19 assembly) using TopHat (version 1.1.3).[14] Only uniquely mapped reads were subjected to downstream analyses. A non-redundant hg19 transcriptome was assembled from UCSC RefSeq genes and UCSC genes. Gene expression in the form of RPKM was calculated using a self-developed script. The search for enriched peaks was performed by scanning each gene using 100-nucleotide sliding windows, and calculating an enrichment score for each sliding window.[15] Windows with RPKM ≥ 5 in the eluate, enrichment score ≥ 2 in genes with RPKM in the input sample ≥1 were defined as enriched in m6A pull down. Enriched windows with score greater than neighboring windows were selected as m6A peaks. To determine "high-confidence" sites, we intersected the peaks with m^6A sites from ref 16.

PARS mapping of RNAs. Nuclease V1 and S1 were used to map secondary structure of RNAs as described.[10] PARS scores for the neighborhood of RRACH motifs were obtained from ref 10.

Table S1. MALDI-MS data for synthetic oligonucleotides in this study

Experiment	Sequence	Mass calculated	Mass found
Melting Set 1	UAC GAC UG	2509.6	2509.3
	UAC G(m^6A)C UG	2523.6	2523.3
	CAG ACG UA	2532.6	2532.0
	CAG GCG UA	2548.6	2548.1
	CAG CCG UA	2508.6	2507.2
	CAG UCG UA	2509.6	2507.6
Melting Set 2	CGA UAG GU	2549.6	2547.3
	CGA U(m^6A)G GU	2563.6	2561.5
	ACC UAU CG	2469.5	2468.8
Dangling Ends	CGCGCG	1889.2	1888.5
	(m^6A)CGCGCG	2232.4	2231.0
	CGCGCG(m^6A)	2232.4	2231.2
	ACGCGCG	2218.4	2217.1
	CGCGCGA	2218.4	2216.2
NMR	GGA CUA GUC C	3159.9	3159.6
	GG(m^6A) CUA GUC C	3173.9	3173.9

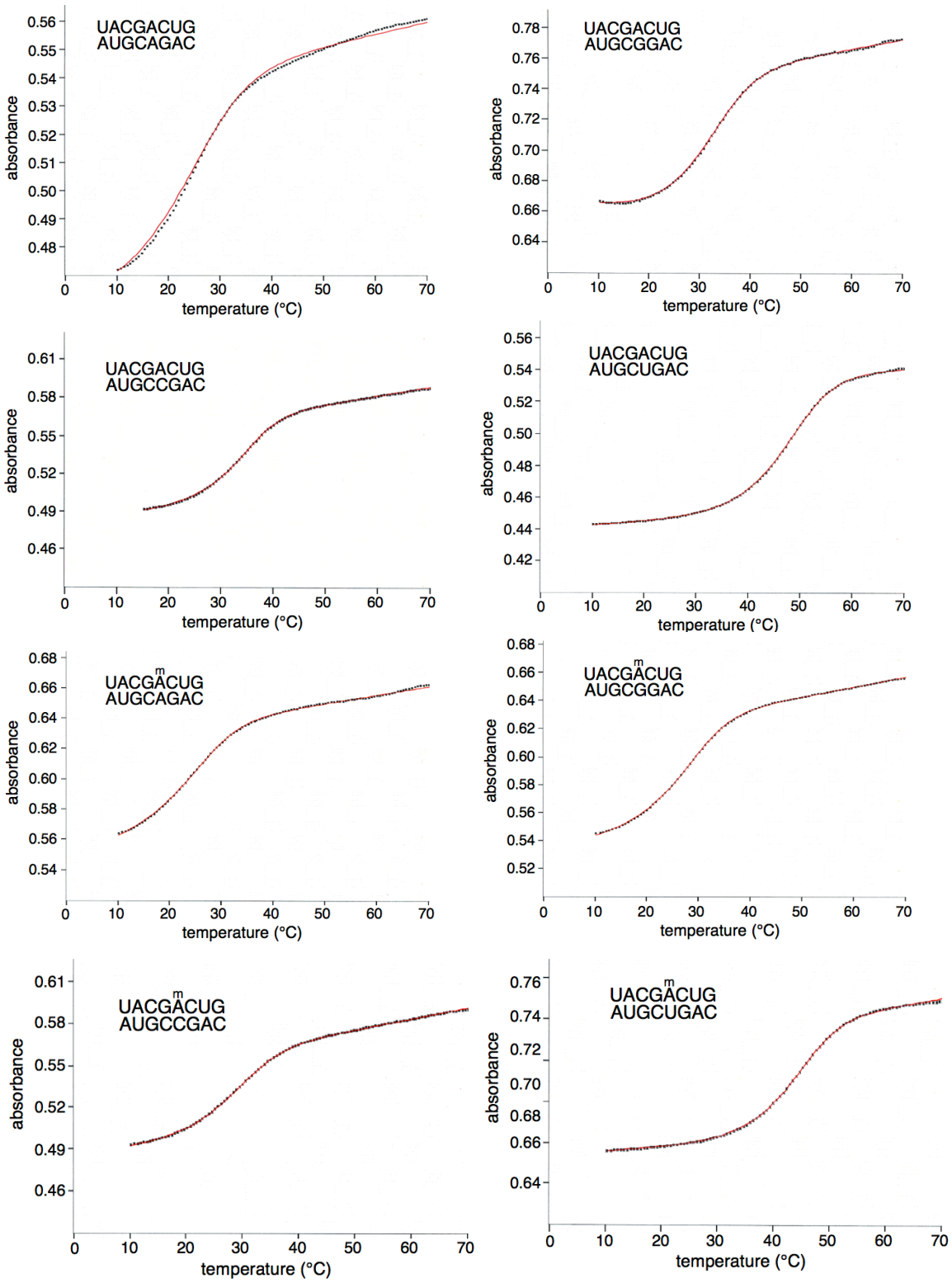


Figure S1. Representative thermal denaturation plots for duplexes in this study. Sequences are shown in each plot. Black points represent measured data; red lines are fits to the two-state model. See "Thermal Denaturation Studies" above for conditions.

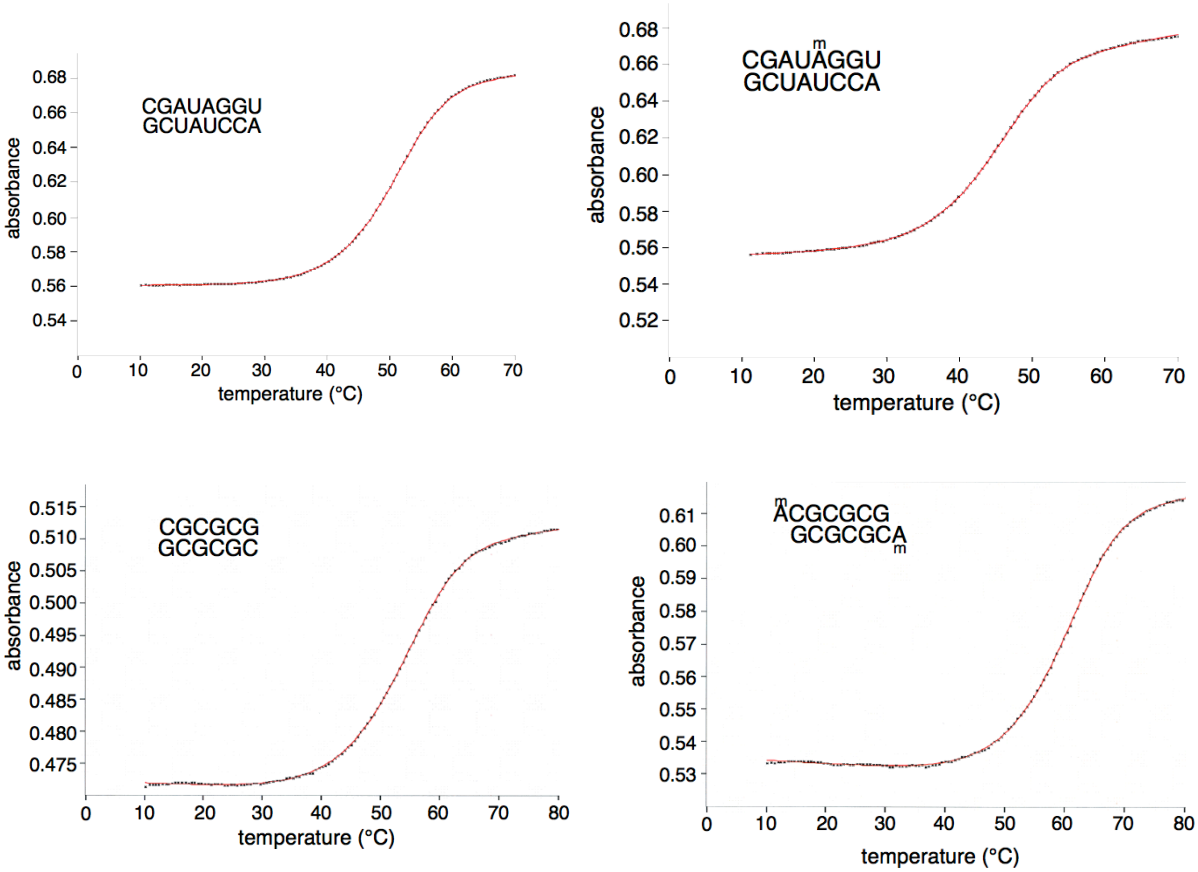
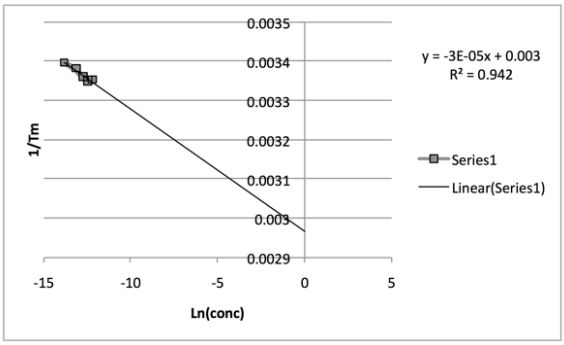
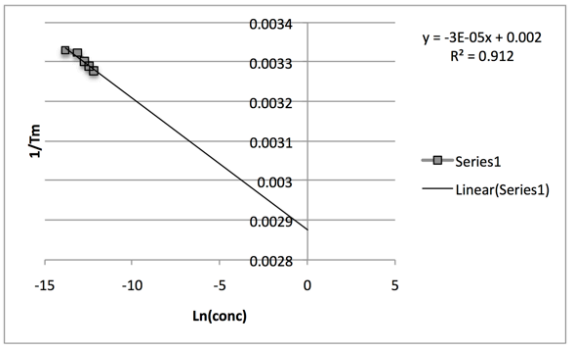


Figure S1 (contd). Representative thermal denaturation plots for duplexes in this study. Sequences are shown in each plot. Black points represent measured data; red lines are fits to the two-state model. See "Thermal Denaturation Studies" above for conditions.

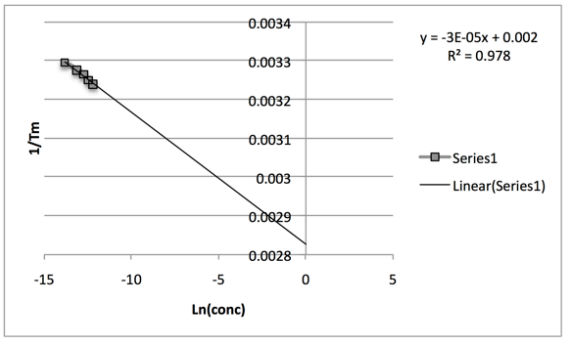
UAC GAC UG + CAG ACG UA



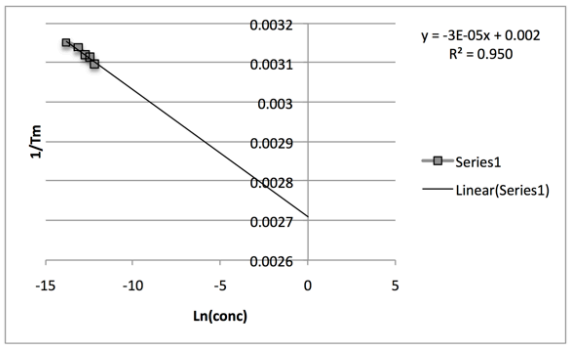
UAC GAC UG + CAG GCG UA



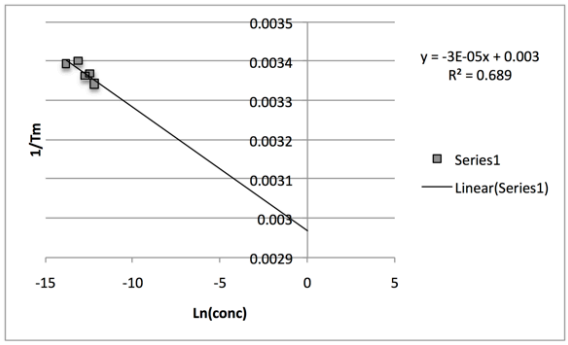
UAC GAC UG + CAG CCG UA



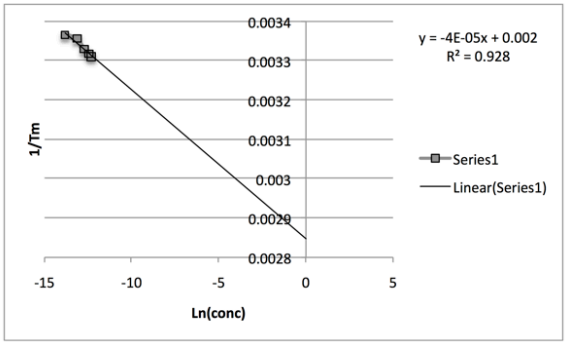
UAC GAC UG + CAG UCG UA



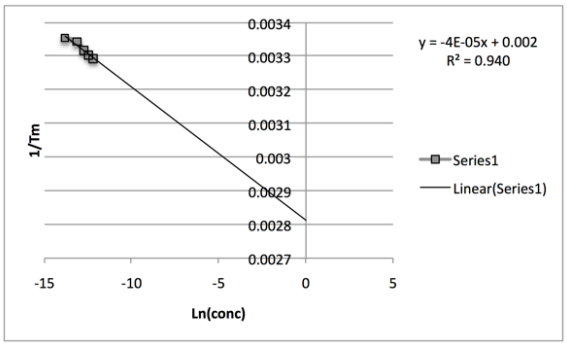
UAC G(m^{6}A)C UG + CAG ACG UA



UAC G(m^{6}A)C UG + CAG GCG UA



UAC G(m^{6}A)C UG + CAG CCG UA



UAC G(m^{6}A)C UG + CAG UCG UA

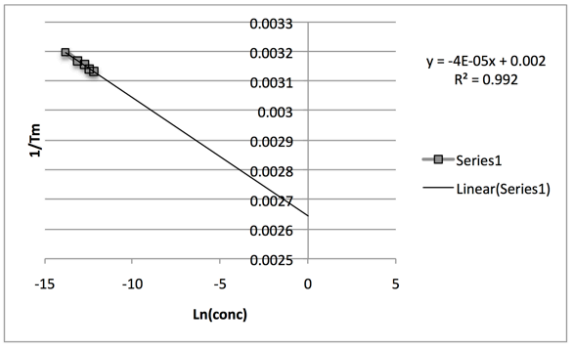
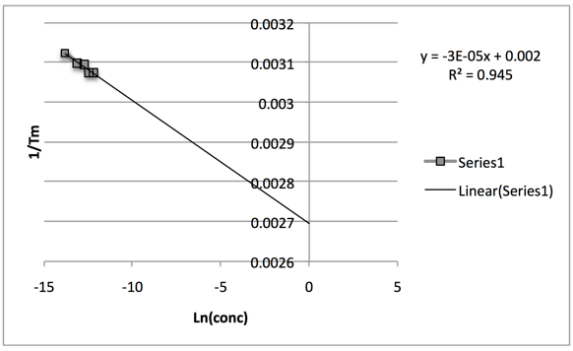


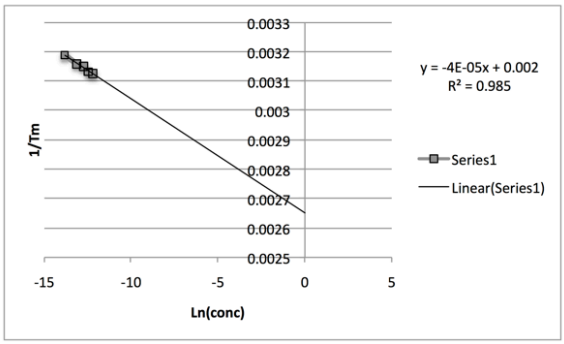
Figure S2. Representative van't Hoff plots for RNA duplexes in this study. See

"Denaturation Methods" above for details. Conditions: 10 mM sodium phosphate buffer, pH 7.0 with 1 M NaCl.

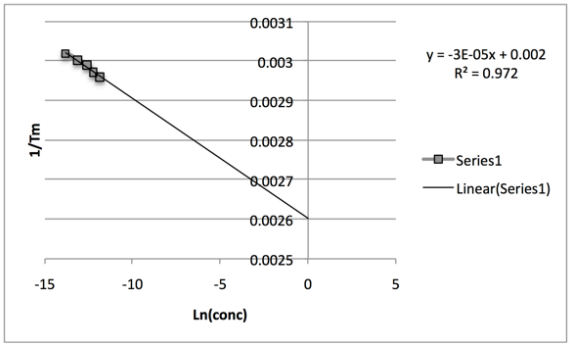
CGA UAG GU + ACC UAU CG



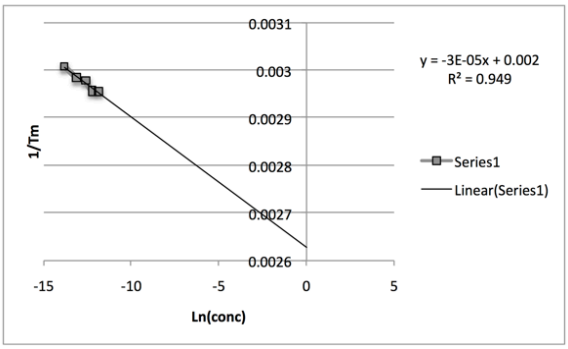
CGA U(m^6A)G GU + ACC UAU CG



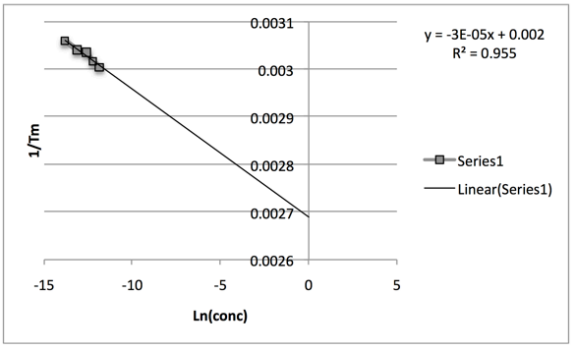
CGC GCG A



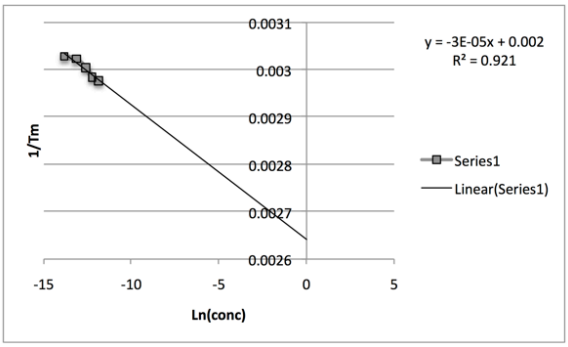
CGC GCG (m^6A)



ACG CGC G



(m^6A)CG CGC G



CGC GCG

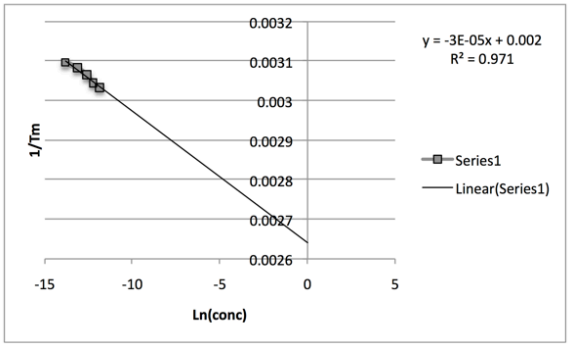


Figure S2 (contd). Representative van't Hoff plots for RNA duplexes in this study. See "Thermal Denaturation Methods" above for details. Conditions: 10 mM sodium phosphate buffer, pH 7.0 with 1 M NaCl.

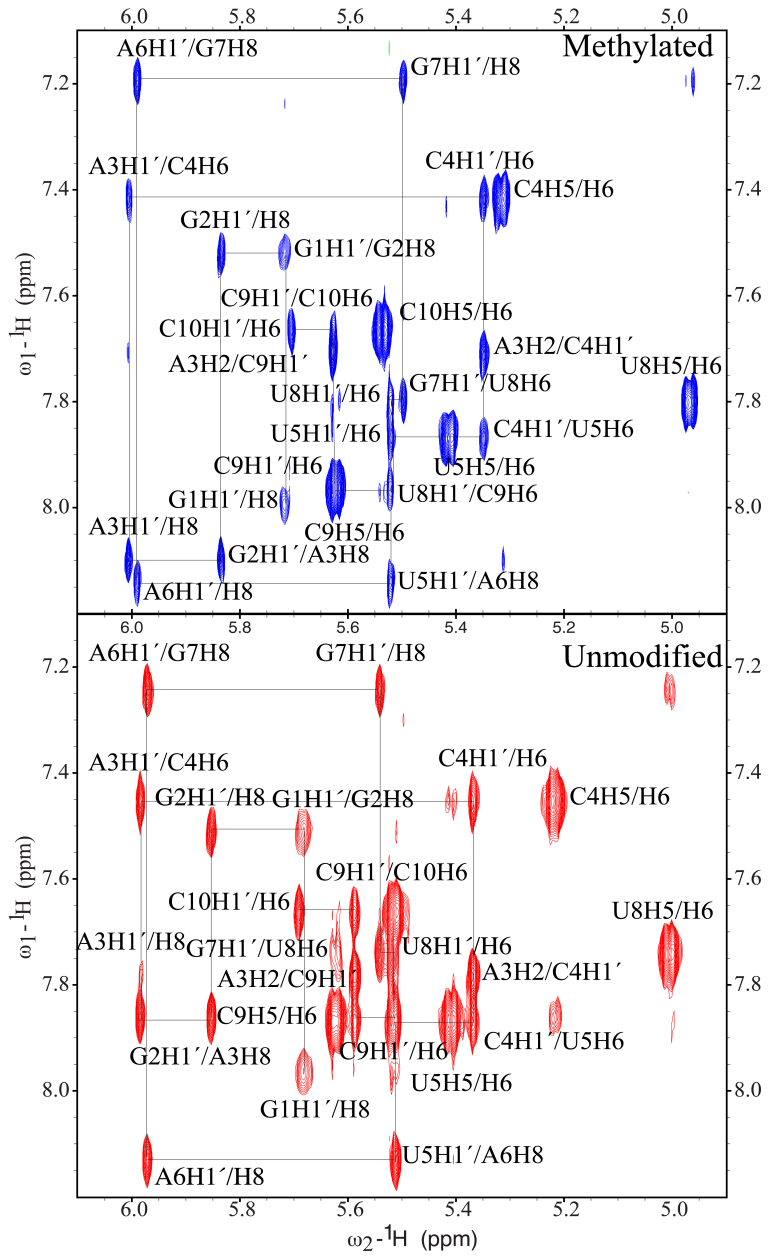


Figure S3. The H1'/aromatic portion of a 150ms mixing time NOESY in D_2O is shown for methylated MA RNA (top, blue) and unmethylated AD RNA (bottom, red), highlighting the sequential assignments of each oligonucleotide.

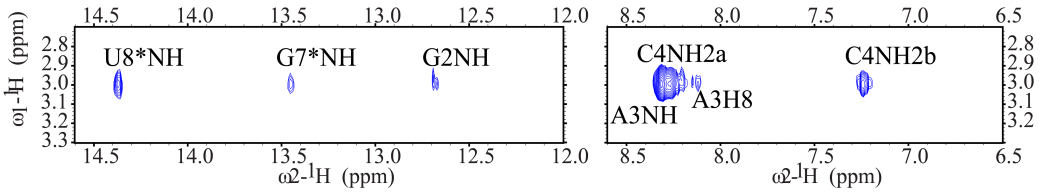


Figure S4. A portion of a 100ms mixing time SS-NOESY experiment acquired in H_2O on MA RNA is shown, highlighting the NOEs from the methyl group of N-methyl adenosine. Nucleotides on the opposing strand of the symmetric duplex are indicated with an *.

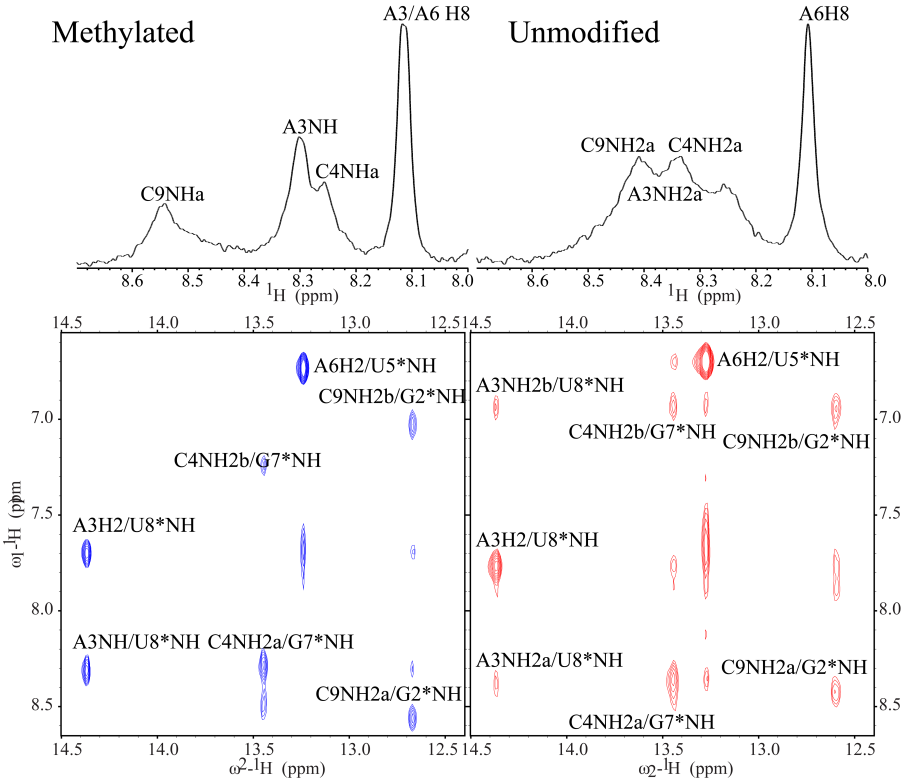


Figure S5. A comparison of the amino group of adenosine with that of N-methyl adenosine is shown. On top, the amino region of the 1D ^{1}H with 11 solvent suppression in 90%H_2O/10%D_2O is shown for MA RNA (left) and AD RNA (right). On bottom, the amino/imino portion of the 250ms SS-NOESY is shown for both sequences MA RNA (left, blue) and AD RNA (right, red), highlighting the relative sharpness of the NH for N-methyl adenosine. Nucleotides on the opposing strand of the symmetric duplex are indicated with an *.

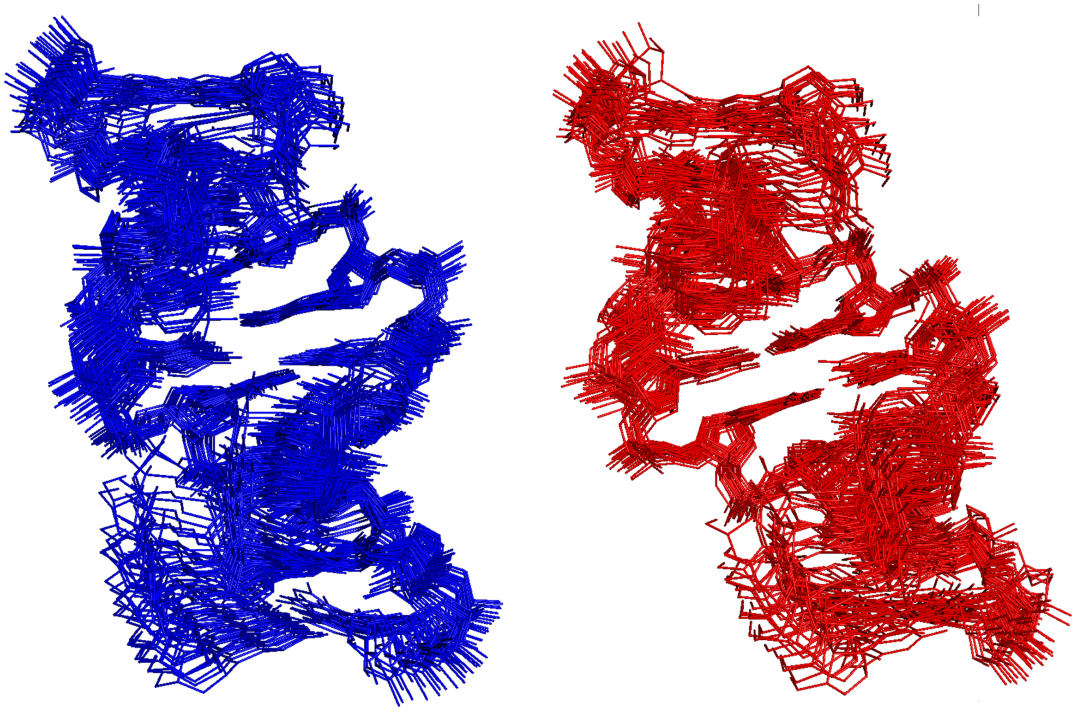


Figure S6. A comparison of the superposition of 25 structures is shown for both oligonucleotides. The methylated RNA is shown in blue, and unmodified RNA is shown in red. The r.m.s.d. values for 25 structures were 0.87 Å and 0.96 Å respectively.

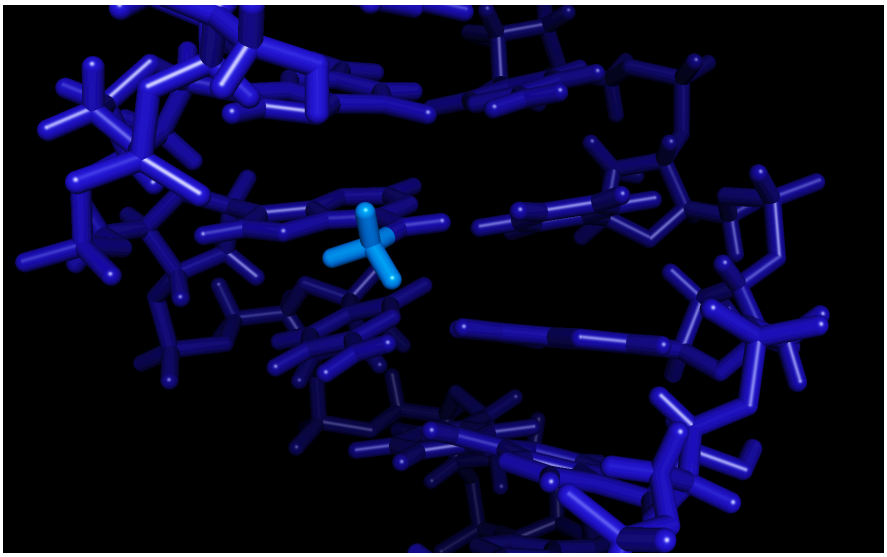


Figure S7. View of m6A looking into the major groove of the methylated MA duplex. The methyl group is light blue, while the rest of the RNA is dark blue.

References

1. Harcourt, E. M.; Ehrenschwender, T.; Batista, P. J.; Chang, H. Y.; Kool, E. T. Identification of a selective polymerase enables detection of N(6)-methyladenosine in RNA. *J. Am. Chem. Soc.* **2013**, *135*, 19079-82.
2. Goddard, T. G.; Huang, C. C.; Ferrin, T. E. *Structure* **2005**, *13*, 473-482.
3. Smallcombe S., Solvent suppression with symmetrically-shifted pulses, *J. Am. Chem. Soc.* **1993**, *115*, 4776-4785.
4. Schwieters, C. D.; Kuszewski, J. J.; Tjandra N.; Clore, G. M. The Xplor-NIH NMR Molecular Structure Determination Package, *J. Magn. Res.* **2003**, *160*, 66-74.
5. Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M., Using Xplor-NIH for NMR molecular structure determination, *Progr. NMR Spectroscopy* **2006**, *48*, 47-62.
6. Wimberly, B.; Varani, G.; Tinoco Jr, I. The conformation of loop E of eukaryotic 5 S ribosomal RNA, *Biochemistry* **1993**, *32*, 1078–1087.
7. Saenger, W. Principles of Nucleic Acid Structure, Springer-Verlag, Berlin (1984).
8. Marino, J. P.; Schwalbe, H.; Griesinger, C. J-coupling restraints in RNA structure determination, *Accts. Chem. Res.* **1999**, *32*, 614–623.
9. Sklenár, V.; Miyashiro, H.; Zon, G.; Miles, H. T.; Bax, A. Assignment of the ^{31}P and ^{1}H resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Lett.* **1986**, *208*, 94-8.
10. Wan, Y.; Qu, K.; Zhang, Q. C.; Flynn, R. A.; Manor, O.; Ouyang, Z.; Zhang, J.; Spitale, R. C.; Snyder, M. P.; Segal, E.; Chang, H. Y. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **2014**, 505, 706-9.
11. Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., Carter, A. C., Flynn, R. A., Zhou, C., Lim, K., Dedon, P., Wernig, M., Mullen, A. C., Xing, Yi., Giallourakis, C. C., Chang, H. Y. m6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*. **2014**, *15*, 707-719.
12. Xiao, R.; Moore, D.D. DamIP: using mutant DNA adenine methyltransferase to study DNA-protein interactions in vivo. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*. **2011**, Chapter 21, Unit21 21
13. Levin, J. Z.; Yassour, M.; Adiconis, X.; Nusbaum, C.; Thompson, D. A.; Friedman, N.; Gnirke, A.; Regev, A. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods* **2010**, 7, 709-15.
14. Trapnell, C.; Pachter, L.; Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **2009**, 25, 1105-11.
15. Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; Sorek,

R.; Rechavi, G. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. **2012**, 485, 201-6.

16. Schwartz, S.; Agarwala, S. D.; Mumbach, M. R.; Jovanovic, M.; Mertins, P.; Shishkin, A.; Tabach, Y.; Mikkelsen, T. S.; Satija, R.; Ruvkun, G.; Carr, S. A.; Lander, E. S.; Fink, G. R.; Regev, A. Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5' Sites. *Cell Reports* **2014**, 8, 284-96.