Supporting Information

Atomistic and thermodynamic analysis of the N6-methyladenosine (m^6A) recognition by the reader domain of YTHDC1

Yaozong Li[a,b‡], Rajiv Kumar Bedi[a‡], Lars Wiedmer[a], Xianqiang Sun[c], Danzhi Huang[a], Amedeo Caflisch[a*]

[a]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

[b]Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden

[c]Regor Pharmaceuticals, Inc., 1206 Zhangjiang Rd, Building C, Pudong New District, Shanghai 201210, China

[*]To whom correspondence should be addressed. Tel: +41 44 635 5521; Email: caflisch@bioc.uzh.ch

‡ These authors contribute equally to this work.

MATERIALS AND METHODS

Cloning, protein expression and purification

The plasmid expressing the N-terminally hexahistidine-tagged YTH domain (residues 345−509) of the human YTHDC1 protein was obtained as a gift from Cheryl Arrowsmith (Addgene ID: 64652) which was also used to generate hexahistidine-tagged T379V mutant using primers listed below. The methods of protein expression and purification were reported by Xu *et al.*[1]

GST_TEV_YTHDC1 (YTH domain residues 345-509) was cloned into pGEX-6P-1 vector between BamHI-XhoI site, amplified using primers GST_TEV_DC1_F and GST_TEV_DC1_R. Single point YTHDC1 M438A and S378A mutants were generated using primers listed below. The protein was overexpressed for 16 hours at 20°C in *Escherichia coli* BL21 (DE3) cells upon induction with 0.2 mM IPTG. The cells were harvested and resuspended in the lysis buffer containing 100 mM Tris-HCl at pH 8.0 and 500 mM NaCl. After lysing the cells by sonication, the cell lysate was clarified by centrifugation at 48000 g for two hours and the supernatant was mixed Glutathione Sepharose 4B (GE Healthcare, 17-0756-01). After extensive washing with the lysis buffer the GST_TEV_YTHDC1 protein was eluted with elution buffer containing 100 mM Tris-HCl at pH 8.0, 500 mM NaCl and 10 mM reduced glutathione. The N-terminal GST-tag was removed by cleavage with tobacco etch virus (TEV) protease at 1:50 ratio. The mixture containing cleaved GST, TEV, and YTHDC1 was finally subjected to a gel filtration step using Superdex 75 16/60 column in a buffer containing 20 mM Tris-HCl at pH 8 and 150 mM NaCl. The protein was concentrated to 10 mg/mL, flash-frozen in liquid nitrogen and stored at -80 °C for future experiments.

List of primers used

T379VF: 5’-GTGTATGGTCCGTGCTCCCTGTAAATGAG-3’

T379VR: 5’-CTTCCCACATACCAGGCACGAGGGACATT-3’

GST_TEV_DC1_F:

5’-ATATGGATCCGAGAACCTGTATTTTCAAGGCACCAGTAAACTCAAATATGTG-3’

GST_TEV_DC1_R:

5’-ATGCCTCGAGTTAGTGACGCATTTTATGAATGACCTGATACAAGTCAATACTTTCATC-3’

M438AF: 5’-GGAATGAGTGCTAAAGCGCTGGGAGGTG-3’

M438AR: 5’-ACTCACGATTTCGCGACCCTCCACAGAA-3’

S378AF: 5’-CGAAGGGTGTATGGGCCACGCTCCCTGTAAATGAG-3’

S378AR: 5’-GGTTTCGCTTCCCACATACCCGGTGCGAGGGACAT-3’

Crystallography

The crystals of YTHDC1 YTH domain (GST_TEV_YTHDC1 apo and mutants) were obtained by mixing 1 µL protein solution at 10 mg/mL with mother liquor containing 0.1 M Bis-Tris at pH 6.5, 0.2 M ammonium sulfate, and 25% PEG 3350 at 22°C in a hanging drop vapor diffusion setup. To obtain crystals of protein complexed with m6A, the crystals were transferred to a 1 µL drop containing 100 mM m^6A (Combi-Blocks, QB-1055) directly dissolved in 0.1 M Bis-Tris at pH 6.5, 0.2 M ammonium sulfate, and 30% PEG 3350, soaked overnight at 22 °C, harvested and frozen in liquid nitrogen without additional cryoprotection.

Diffraction data were collected at the Swiss Light Source (Villigen, Switzerland) using the beamline X06DA (PXIII), and processed using XDS.[2] The structures were solved by molecular replacement using Phaser program[3] from the Phenix package.[4] The unliganded structure of YTHDC1 (PDB ID: 4R3H) was used as a search model. The model building and refinements were performed using COOT[5] and phenix.refine.[4] Data collection and refinement statistics are summarized in Table S1.

Coordinates

The atomic coordinates and structure factors for the 11 crystal structures have been deposited in the Protein Data Bank (PDB) with the accession codes 6ZCN, 6ZD9, 6ZD3, 6ZDA, 6ZD4, 6ZD5, 6ZD8, 6ZD7, 6YNN, 6YNO, and 6ZCM.

Binding assay

Isothermal titration calorimetry (ITC) experiments were carried out at 18˚C using MicroCal ITC200 (GE Healthcare). All m6A experiments were performed in gel filtration buffer (20 mM Tris-HCl at pH 8 and 150 mM NaCl), while all GGm6ACU experiments were performed in 20 mM HEPES, pH 7.5, 150 mM NaCl. m6A was directly dissolved at 10 mM in the buffer and titrated into the sample cell containing the protein at 1 mM. GGm6ACU was dissolved at 500 µM in the buffer and titrated into the sample cell containing the protein at 50 µM. After an initial injection of 0.4 µL, 19 injections of 2.0 µL each were performed. The raw data were integrated, normalized for concentration, and analysed using a single binding site model, provided in the MicroCal Origin software package.

Differential scanning fluorimetry (DSF)

Differential scanning fluorimetry (DSF) is a method to monitor the thermal unfolding of a protein using a hydrophobic binding dye like SYPRO® orange. The DSF measurements for the mutants and wildtype were carried out with a final volume of 20 µL on a LightCycler® 480 II Real-Time PCR machine (Roche). The assay buffer contained 50 mM HEPES at pH 7.5 and 150 mM NaCl in which each protein was diluted to a final concentration of 2 µM. The SYPRO® orange was used as a dilution of 1:1000 of the manufacturer's stock. For the thermal denaturation, the temperature was increased from 25 °C to 90 °C by a heating rate of 3.2 °C/min. The change in fluorescence upon protein unfolding was consecutively monitored by exciting at 465 nm and recording the emission at 590 nm during heating. The protein melting curves were plotted as a derivative function (- dF/dt) using the Tm Calling Analysis module of LightCycler® 480 BasicSoftware. The melting temperature (Tm) was the temperature giving the minima of the derivative melting curve, which also represents the inflection point of the melting curve.

Reproducing the positions of water 1 in its crystal structures by SZMAP:

The SZMAP method analyzes the binding site of a targeted protein with or without its ligands using the semi-continuum solvation theory. The hybrid solvation theory combines a single explicit probe water molecule with the Poisson-Boltzmann continuum theory. The probe water is sampled extensively in different orientations at preset grids in the binding site, and its interactions with continuum solvent and with protein and ligand molecules are evaluated accordingly. Thereby, the thermodynamic properties about the probe water bound to the binding site are estimated. A key point that can be predicted by SZMAP is the thermodynamical stability of a water molecule at a specific region of the binding site. Thereby, stable, and unstable waters can be identified.

The complex structure of YTHDC1-m^{6}A (PDBid: 6ZCN) was prepared as the procedure for CHARMM and NAMD simulations. The atom names of the minimized structure were adapted for the force field AmberFF94 and AM1-BCC partial charges were assigned in the B-factor column of the structure. The stabilization calculation was then carried out by SZMAP. The calculation automatically conducted separate calculations on apo and holo structures. The free energy difference of the neutral probe was used to examine the stability of predicted water. The values for water 1 are listed in **Table S4**.

Binding free energy calculations by inhomogeneous solvation theory (IST):

Four systems were simulated and analyzed by the IST protocol, including apo and holo (bound to m^{6}A) proteins in both wild type and Thr379Val mutant. The systems were prepared the same as mentioned above. However, MD simulations were carried out using Desmond (Schrodinger 2018-4 Release) with the OPLS2005 force field for further IST analysis. The system was firstly minimized for 2000 steps using a steep descent algorithm with 1 kcal/mol/Å as the convergence threshold and then was equilibrated with the default protocol implemented in Desmond. In the production MD simulation, Nosé-Hoover chain thermostat[6] was applied to control the system temperature at 300 K and Martyna-Tobias-Klein Barostat[7] was used to control the system pressure at 1 atm. The cutoff values were set to 9 Å for short-range electrostatic and van der Waals interactions. The time step was set to 2 fs. To maximum the conformational sampling of the water molecules in the protein-ligand interface or ligand-

binding pocket, the protein was restrained with a force of 15 kcal/mol/Å on the backbone and 5 kcal/mol/Å on side-chain atoms. Such restraint was also applied in our previous work and showed an efficient sampling of the interested water molecules.[8] Each of the four systems was simulated for 20 ns and 10,000 snapshots were collected for the later IST analysis.

Locating hydration sites:

A stepwise protocol was used to determine water sites where water molecules will occupy. First, a spherical boundary in a radius of 8 Å was defined to cover the whole interested region which was centered by the geometrical center of m^{6}A, and the sphere was gridded by cubes in the length of 1.2 Å. Then, the water molecules in each frame of an MD trajectory were assigned to the grids. The grids occupied by water molecules in more than 20% of the simulation time were defined as initial water sites. Finally, iterations were carried out by calculating the geometric center of the water molecules in the initial water sites along the trajectory until the center drift was less than 0.01 Å or 20 times of iteration was reached. During this process, water molecules within 1.3 Å of the geometric center were assumed to be in the same water site. Two centers were merged if their distance is less than 1.0 Å.

Solvation free energy:

ΔG_s describes the change of the solvation energy upon the relocation of one water molecule from the water site to bulk water,

$$\Delta G_s = \langle E_{hs} \rangle - \langle E_{bulk} \rangle - (TS_{hs} - TS_{bulk}) - P\Delta V \quad (1)$$

Herein, $\langle E_{hs} \rangle$ is the averaged interaction energy between the water molecule in the water site and the rest of the system. $\langle E_{bulk} \rangle$ represents the averaged interaction energy between one bulk water with the rest of the system. T is the Kelvin temperature; S_{hs} and S_{bulk} are the entropic contributions to the solvation energy of the water molecule in the water site and the bulk, respectively. The solvation entropy of the water site S_{hs} is separated as,

$$S_{hs} = S_{wp} + S_{ww} \quad (2)$$

where S_{wp} is the water-protein term and S_{ww} water-water reorganization term. The water-protein term is calculated as,

$$S_{wp} = -k\frac{\rho}{\Omega}\int g_{wp}(\boldsymbol{r},\omega) ln g_{wp}(\boldsymbol{r},\omega) d\boldsymbol{r} d\omega \quad (3)$$

where k is the Boltzmann constant, ρ is the density of bulk water, Ω equals $8\pi^2$, $g_{wp}(\boldsymbol{r},\omega)$ is the combined translational and rotational radial distribution function (RDF) of the water molecule. The water-water reorganization entropy is separated into the translational term (S_{ww}^{tr}) and the orientational term (S_{ww}^{or}):

$$S_{ww} = S_{ww}^{tr} + S_{ww}^{or} \quad (4)$$

With the Kirkwood superposition approximation (KSA) which assumes that the pair correlation function in an inhomogeneous solvent is equal to that in the bulk solvent and is only dependent on the distance and relative orientation between the two solvent molecules, S_{ww}^{tr} and S_{ww}^{or} can be calculated as described in the supplementary information.

Supplementary Table S1. Molecular systems simulated in this study.

State	System	Notes	Simulation Type	Length (ns)	Replica	PDB id
Apo	Canonical Apo	Met438 out of the binding pocket	Conventional MD	1000	5	Chain A of 6ZD9
			Metadynamics	500	5	
	Met438-in Apo	Met438 rotates in the binding pocket	Conventional MD	1000	5	Chain B of 6ZD9
	Canonical Apo	Annihilate water 1	Alchemical transformation	81	1	Chain A of 6ZD9
	T379V		Conventional MD	1000	5	6ZD8
	T379V	Annihilate water 1	Alchemical transformation	81	1	6ZD8
Holo	YTHDC1 with GG(m^6A)CU		Conventional MD	1000	5	4R3I
	YTHDC1 with GGCU		Conventional MD	1000	5	4R3I
	YTHDC1 with m^6A		Conventional MD	1000	5	Chain A of 6ZCN
	YTHDC1 with m^6A	Annihilate water 1	Alchemical transformation	81	1	Chain A of 6ZCN
	T379V YTHDC1 with m^6A		Conventional MD	200	5	6ZD7
	T379V YTHDC1 with m^6A	Annihilate water 1	Alchemical transformation	81	1	6ZD7
Ligand	GG(m^6A)CU		Conventional MD	500	10	4R3I
	GGCU		Conventional MD	500	10	4R3I
Hybrid	m^6A → adenosine in YTHDC1		Alchemical transformation	99	1	Chain A of 6ZCN
	m^6A → adenosine in aqueous solution		Alchemical transformation	99	1	-
Bulk water	water 1	Annihilate water 1	Alchemical transformation	66	1	-

Supplementary Table S2. Data collection and structure refinement statistics of YTHDC1 structures.

PDB ID:	6ZCN	6ZD9	6ZD3	6ZDA
Protein	**DC1-m6A**	**GST-DC1**	**M438A**	**M438A-m6A**
Data Collection				
Space group	$P12_11$	$P12_11$	$P12_11$	$P12_11$
Cell dimension a, b, c (Å)	39.69, 106.48, 42.05	39.88, 103.98, 41.98	39.71, 103.13, 42.31	40.09, 103.33, 42.35
Cell dimension α, β, γ (°)	90, 106.48, 90	90, 103.98, 90	90, 105.42, 90	90, 105.92, 90
Resolution (Å)	40.32-1.6 (1.7-1.6)	40.67-1.51 (1.6-1.51)	40.79-1.25 (1.33-1.25)	40.73-1.3 (1.38-1.3)
Unique reflections*	42400 (6707)	51254 (8258)	88036 (13842)	80684 (12941)
Completeness*	99 (97.3)	98.3 (98.1)	97.3 (94.4)	99.7 (99)
Redundancy*	3.23 (3.03)	3.43 (3.47)	4.55 (4.45)	4.48 (4.29)
R_{merge} * %	4.6 (68.1)	3.3 (18.8)	6.8 (134.6)	4.3 (119.5)
CC (1/2)	99.9 (81.6)	99.9 (97.4)	99.9 (51.9)	99.9 (62.9)
I/σI	15.05 (1.72)	23.16 (6.02)	11.71 (1.14)	15.78 (1.35)
Refinement				
R_{work}/R_{free}	19.08/22.48	18.24/20.49	19.21/21.24	20.41/21.96
RMSD bond (Å)	0.005	0.005	0.006	0.005
RMSD angle (°)	0.774	0.788	0.81	0.772
B-factor (Å^2)**	28.88/41.72/37.57	25.47/43.59/35.22	18.61/35.73/31.19	24.7/30.37/33.54
Ramanchandran Favored (%)	99.38	99.07	99.34	98.77
Ramanchandran allowed (%)	0.62	0.93	0.66	1.23
Ramanchandran Disallowed (%)	0	0	0	0

PDB ID:	6ZD4	6ZD5	6ZD8	6ZD7
Protein	**S378A**	**S378A-m6A**	**T379V**	**T379V-m6A**
Data Collection				
Space group	$P12_11$	$P12_11$	$P12_11$	$P12_11$
Cell dimension a, b, c (Å)	39.51, 103.49, 42.03	39.82, 103.4, 42.49	39.76, 103.92, 41.94	39.93, 41.78, 103.85
Cell dimension α, β, γ (°)	90, 105.75, 90	90, 105.76, 90	90, 104.67, 90	90, 90, 74.76
Resolution (Å)	40.49-1.4 (1.49-1.4)	40.9-2.3 (2.44-2.3)	40.58-1.5 (1.59-1.5)	40.42-1.75 (1.86-1.75)
Unique reflections*	61908 (9997)	14545 (2296)	52189 (8385)	32985 (5201)
Completeness*	97.4 (97.8)	98.7 (96.8)	99.6 (99.4)	99.3 (97.7)
Redundancy*	4.42 (4.35)	4.5 (4.7)	3.75 (3.53)	3.82 (3.63)
R_{merge}*	7.2 (116.8)	4.8 (12.3)	5.4 (55.8)	6.8 (113.5)
CC (1/2)	99.9 (72.8)	99.8 (99.3)	99.9 (81.2)	99.9 (64.2)
I/σI	13.63 (1.27)	25.77 (12.16)	15.02 (2.28)	11.49 (1.24)
Refinement				
R_{work}/R_{free}	24.66/27.05	18.6/23.31	18.49/21.53	20.66/23.81
RMSD bond (Å)	0.006	0.008	0.005	0.006
RMSD angle (°)	0.822	0.941	0.788	0.818
B-factor (Å^2)**	19.82/38.19/27.69	22.32/32.56/26.53	22.77/47.6/32.25	36.24/49.95/40.17
Ramanchandran Favored (%)	99.04	97.84	99.04	98.15
Ramanchandran allowed (%)	0.96	2.16	0.96	1.85
Ramanchandran Disallowed (%)	0	0	0	0

PDB ID:	6YNN	6YNO	6ZCM
Protein	**DHUDC1135**	**DHUDC1139**	**DHUDC1180**
Data Collection			
Space group	$P12_11$	$P12_11$	$P12_11$
Cell dimension a, b, c (Å)	39.73, 41.94, 103.52	39.64, 41.93, 103.33	39.58, 42, 103.44
Cell dimension α, β, γ (°)	90, 90, 75.53	90, 90, 75.18	90, 90, 104.73
Resolution (Å)	40.72-1.2 (1.27-1.2)	40.65-1.4 (1.49-1.4)	40.68-1.24 (1.32-1.24)
Unique reflections*	101872 (16196)	63898 (10172)	91479 (14554)
Completeness*	99.3 (97.8)	99.5 (98.7)	98.8 (97.4)
Redundancy*	3.33 (3.16)	3.36 (3.26)	4.44 (4.22)
R_{merge}*	5.7 (89.5)	6.4 (118)	7.7 (152.7)
CC (1/2)	99.9 (61.2)	99.9 (50.3)	99.9 (56.1)
I/σI	12.56 (1.47)	12.76 (1.14)	12.37 (1)
Refinement			
R_{work}/R_{free}	19.3/21.5	19.32/22.48	21.63/24.75
RMSD bond (Å)	0.005	0.005	0.005
RMSD angle (°)	0.781	0.774	0.823
B-factor (Å^2)**	16.43/23.41/28.25	20.17/32.76/30.69	17.62/31.45/27.25
Ramanchandran Favored (%)	99.69	99.38	99.07
Ramanchandran allowed (%)	0.31	0.62	0.93
Ramanchandran Disallowed (%)	0	0	0

*Statistics for the highest resolution shell is shown in parentheses.
** P/L/W indicate protein, ligand/ion and water molecules, respectively.

Supplementary Table S3. Binding free energies of water 1 to YTHDC1 in its apo and holo forms. The values are reported in kcal/mol.

	SZMAP[a]		IST[b]		Alchemical[c]	
	Wild Type	Thr379Val	Wild Type	Thr379Val	Wild Type	Thr379Val
Holo	-11.3	-6.4	-6.9	-2.4	-11.4 ± 0.1[d]	-9.3 ± 0.1
Apo	-10.8	-5.6	-5.1	-1.6	-8.2 ± 0.1	-8.3 ± 0.1
Bulk	0	0	--	--	-6.3 ± 0.1	-6.3 ± 0.1
$\Delta Gw1_{holo}$	--	--	-6.9	-2.4	-5.1 ± 0.1	-3.0 ± 0.1
$\Delta Gw1_{apo}$	--	--	-5.1	-1.6	-1.9 ± 0.1	-2.0 ± 0.1

[a] The neutral difference free energy was reported. In all four systems, water 1 was correctly placed as its position in the corresponding crystal structure. The negative values between holo and apo forms indicate that water 1 is stabilized in its position by the m^{6}A binding.

[b] The binding free energy of water 1 to YTHDC1 was calculated based on IST. water 1 was correctly predicted in its position of the corresponding crystal structures.

[c] The binding free energy of water 1 to YTHDC1 was estimated by alchemical simulations. The double decoupling approach was used to obtain the binding free energy of water 1 for each system. $\Delta Gw1_{holo}$ and $\Delta Gw1_{apo}$ designate the binding free energy for holo and apo systems, respectively.

[d] The statistical error was estimated by the bootstrap method.

Figure S1. Integrands and convergency of the alchemical transformations from m^6A to adenosine in protein and aqueous environment. **a)** and **b)** present the integrands for transforming m^6A to adenosine in protein and aqueous environment, respectively. The error bars shown together with the integrands show 2 folds of the standard deviation of $dG/d\lambda$ and thereby reflect how well the phase overlap between neighbouring λ simulations. The free energy changes for 3 individual steps are shown together with their associated statistical errors. The errors were estimated using the bootstrapping method. Free energies of three alchemical transformation steps, i.e., unchanging, vdW transformation, and charging, are shown in the bottom of the **Figure S1c)** The convergence of free energy difference for transforming m^6A to adenosine. The free energy difference for m^6A to adenosine in protein and bulk water (ΔΔG) are plotted with the time evolution. The first nanosecond of the simulation was considered as the equilibration phase and the latter two nanoseconds of data were used to calculated ΔΔG.

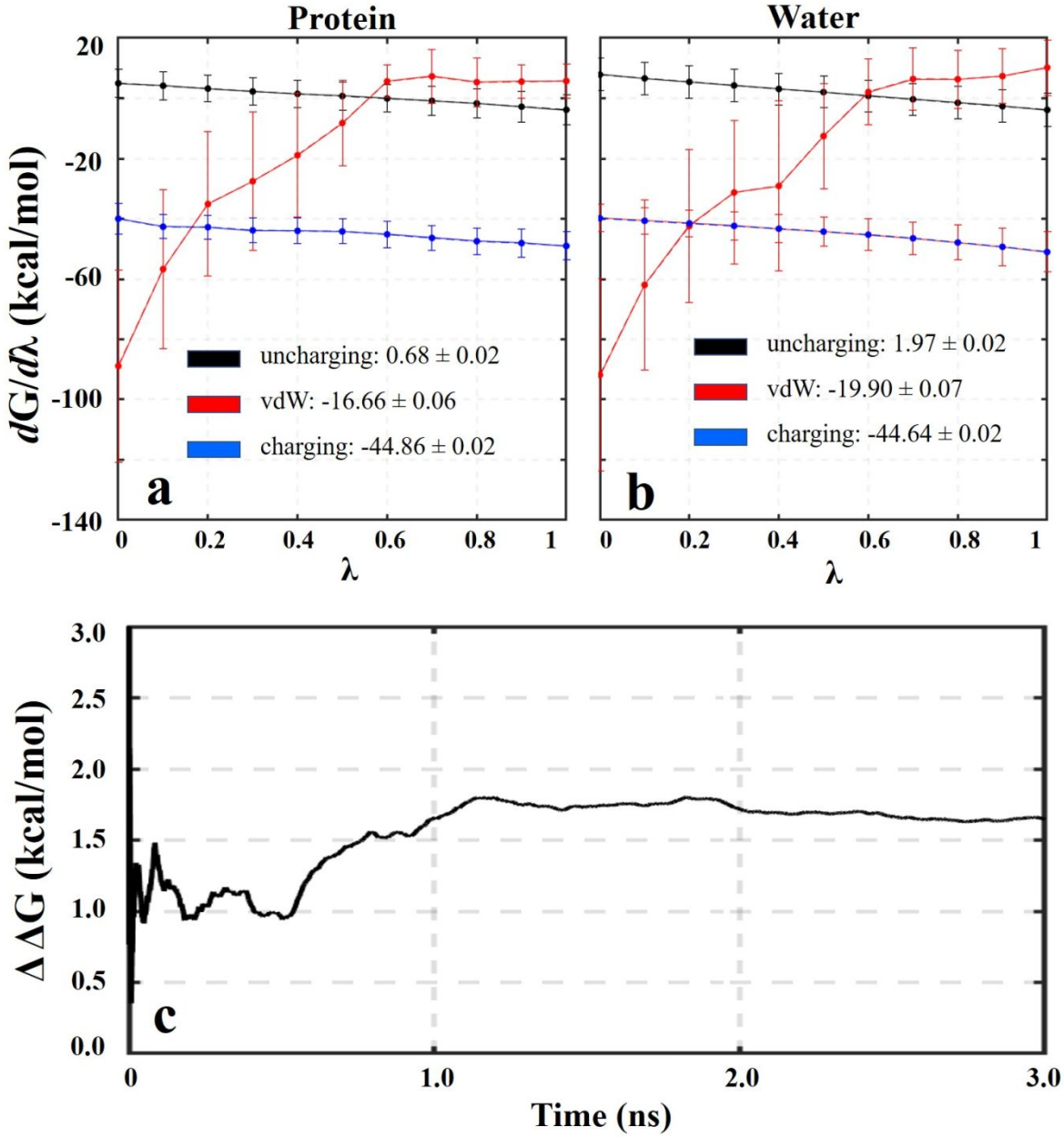


Figure S2. Pose variation of the un-methylated adenosine in the YTHDC1 binding pocket. RMSD values of the binding poses of the (un)methylated adenosine of the penta-oligoribonucleotides (GGACU) with respect to the m^{6}A crystal pose (based on the structure 4R3I) are plotted in black and red lines. Five replicas are distinguished by dashed lines. The corresponding PMF plots are shown on the right panel. The PMF is calculated as $-k_B T \ln\frac{p_i}{p_0}$, where k_B is the Boltzmann factor, T is the temperature (300 K), p_i is the population in bin i, and p_0 is the probability of the most populated bin (with a 0.05 Å bin width). Representative binding poses of the unmethylated adenosine are shown together with YTHDC1 structures in different colours. These binding poses are long-lived during MD simulations. Pose 1 corresponds to the binding mode in the crystal structure. Other nucleotides of the oligomers are not shown.

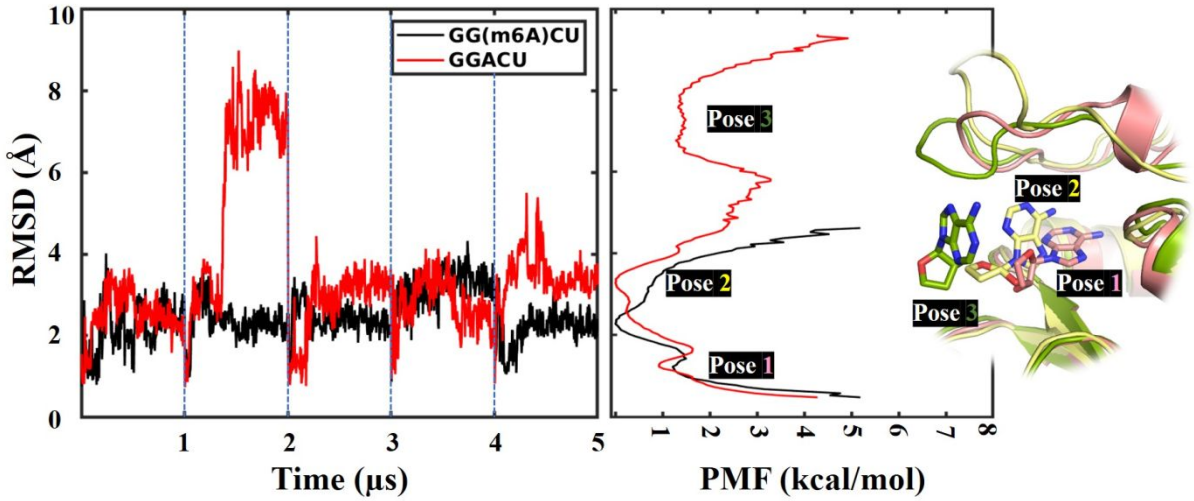


Figure S3. Potential of Mean Force (PMF) for differentiating the methylated and unmethylated pentamer ribonucleotides in the unbound state. **a)**, and **b)** show 2D free energy surface as a function of solvent accessible surface area (SASA) and root mean square deviation (RMSD) for methylated and unmethylated pentamer ribonucleotide GGACU. For calculating SASA values, only the adenine base was considered by excluding the methyl group of m^{6}A. For calculating the RMSD, the AC dinucleotide of the two pentameric systems was superposed to the pose in the crystal structure. The resolution of the energy value is set to 0.5 kcal/mol (legend bar). The dashed circles highlight the region where the bound conformations of the AC dinucleotide are mapped. **c)**, and **d)** show 1D PMF for differentiating the methylated and unmethylated pentamer ribonucleotides in the unbound state. c) PMF along the Root Mean Square Deviation (RMSD) from the reference bound-state structure. The segment (m^{6}A)C in the complex structure (PDB code: 4R3I) is selected as the reference for the bound state, and its heavy atoms are used for the RMSD calculation. MD snapshots of both simulated ribonucleotides in aqueous solution are superposed to the reference segment, and the corresponding RMSD values are calculated. The PMF is calculated as $-k_{\mathrm{B}}T\ln\frac{p_i}{p_0}$, where k_{B} is Boltzmann factor, T is the temperature (300 K), $\mathrm{p_i}$ is the population in bin i, and $\mathrm{p_0}$ is the probability of the most populated bin (with a 0.05 Å bin width). The associated error is estimated using the bootstrapping method. d) Two-state models showing conformational free-energy differences between bound-like and unbound conformations of penta-oligoribonucleotides. The conformational free-energy difference is calculated by the equation $\Delta G = -k_{\mathrm{B}}\,T\ln\frac{p_b}{p_u}$, where k_B is Boltzmann factor, T is the temperature (300 K in this study), p_b is the population of bound-like conformations, and p_u is the population of the unbound conformations ($p_u = 1 - p_b$).

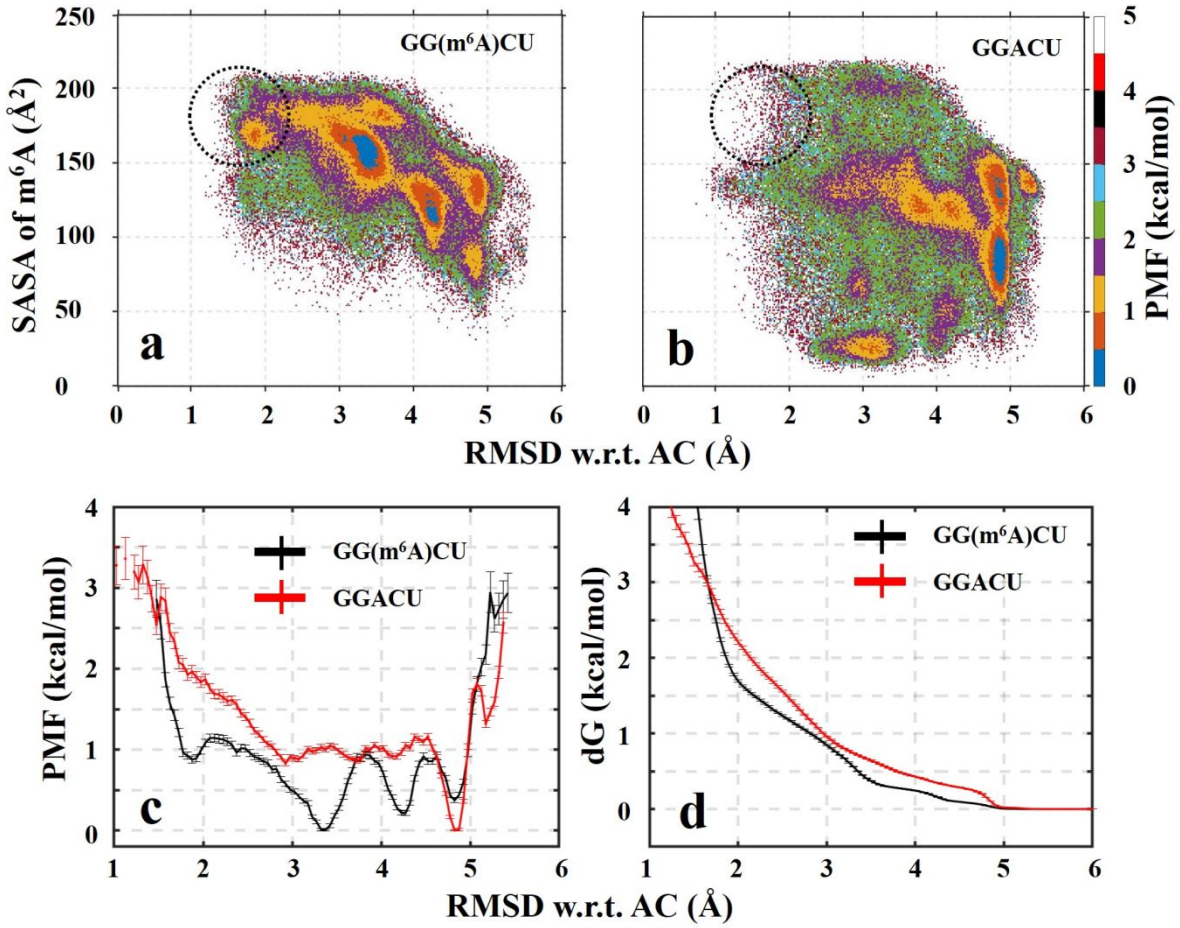


Figure S4. Conformational changes of the residues in the YTHDC1 binding pocket. **a)** The flipping of Trp428. Chain A of the structure 6ZD9 was used for the starting structure of the unbiased MD simulations. **b)** The flipping of Met438. Chain B of the structure 6ZD9 was used for the starting structure of the unbiased MD simulations. Five replicas were produced for each system, and each replica was run for one microsecond (μs). The replica is separated by solid green lines. Representative structures are presented in different colours. The dihedral angles describing the conformational changes are presented on 2D structures.

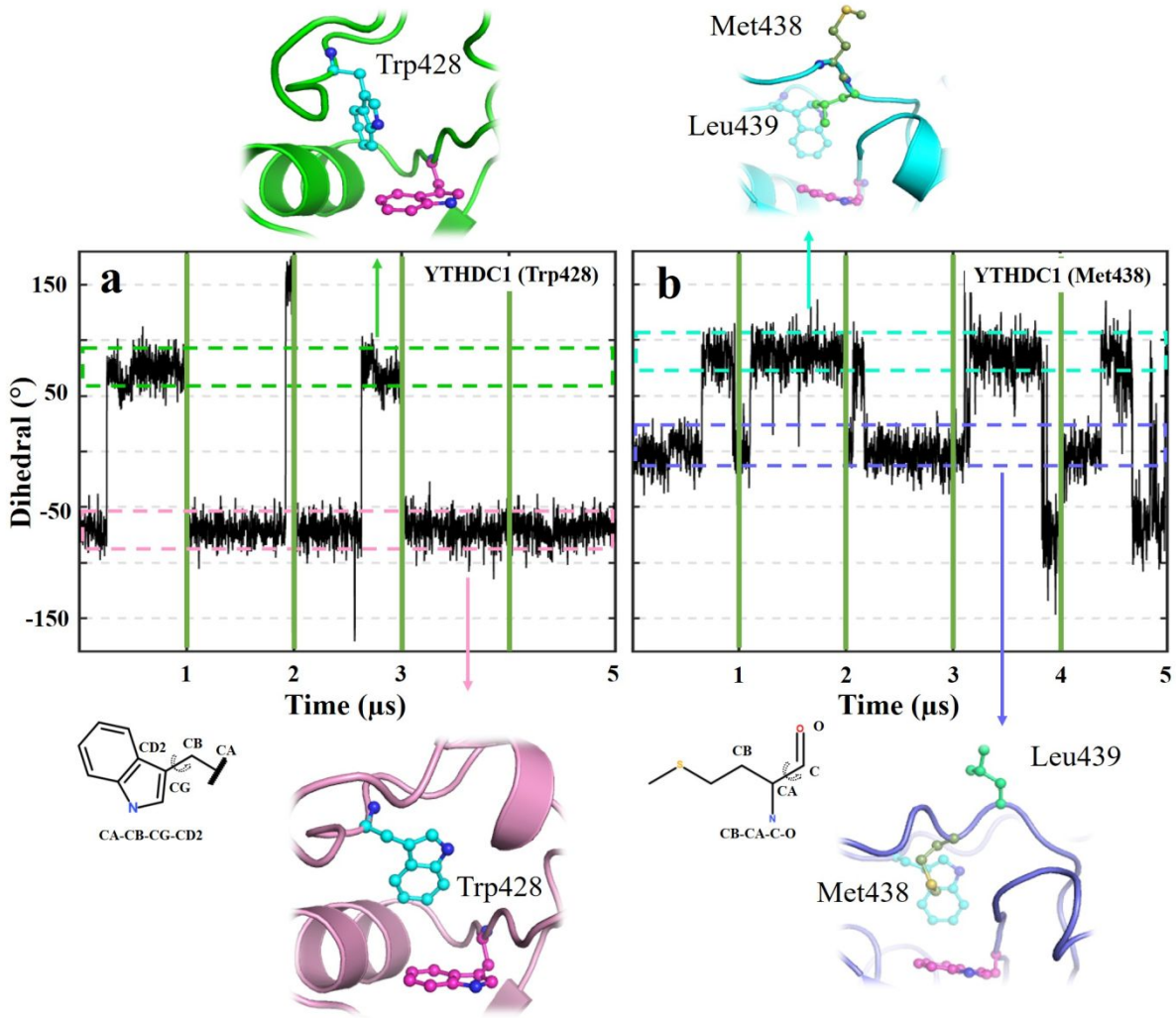


Figure S5. ITC titration curves of binding of m^{6}A nucleoside and GG(m^{6}A)CU.

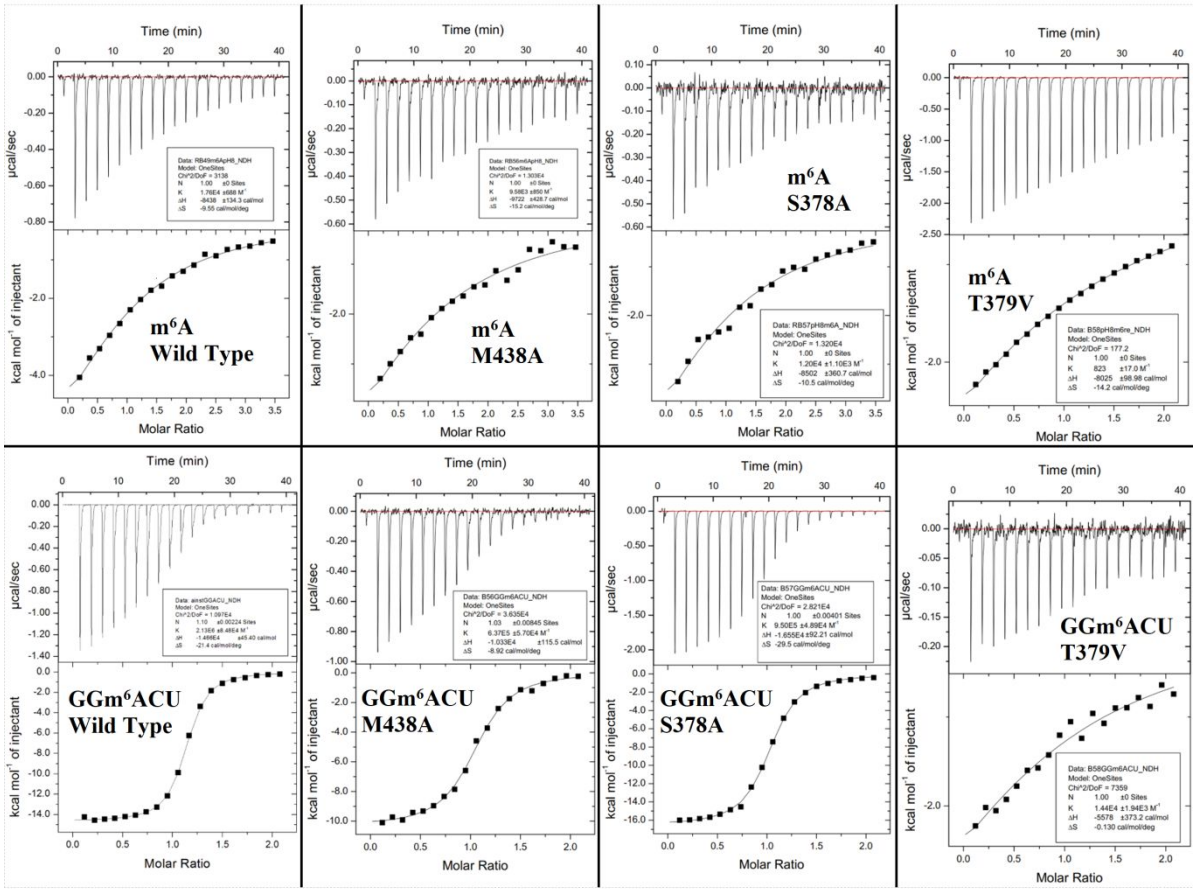


Figure S6. B-factor of different monomers in the structure 6ZD9. The B-factor values of the two monomers were mapped onto the corresponding structures with the same reference value. The thickness and colour of backbone vary smoothly between the thin blue-coloured tube for the smallest B-factor value and the thick red-coloured tube for the greatest B-factor value.

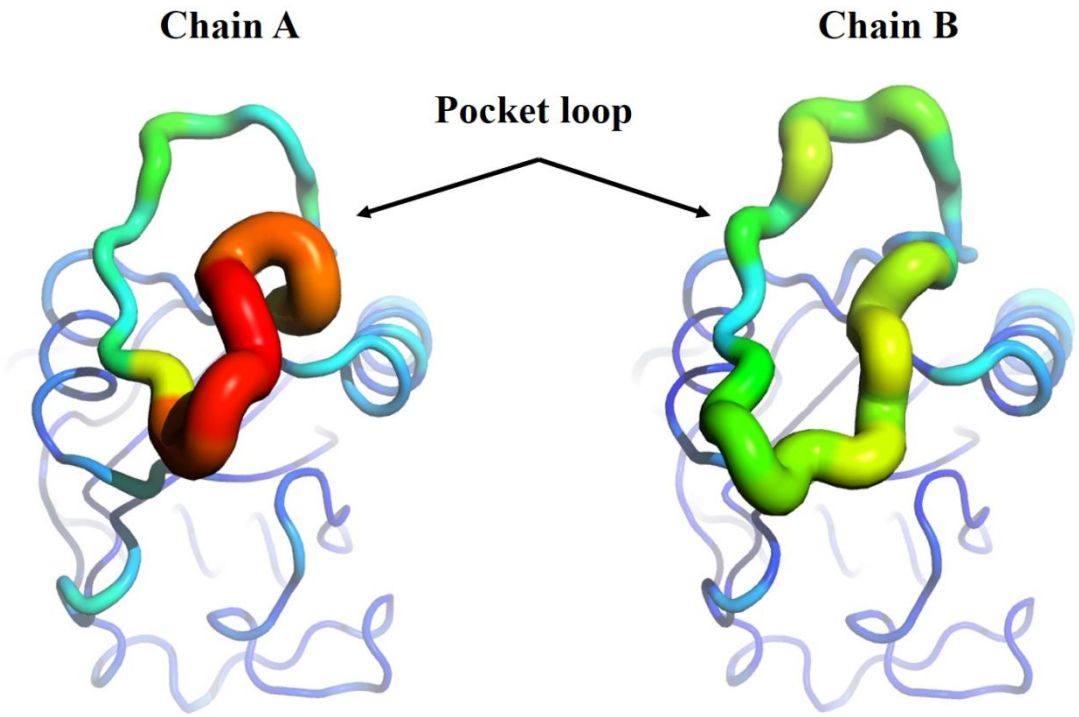


Figure S7. Water distribution of different monomers in the structure 6ZD9. Top panel: The crystal water molecules that are wrapped in the binding pocket and in the middle of the binding pocket loop are shown in the red sphere. The binding loop includes residues 420 to 441 and the binding pocket consists of residues 362 to 364, 367, 377 to 379, 383, 428, 438, 439, and 476. Bottom panel: Population of the number of water molecules sampled by MD simulations for the region mentioned above. Only the first hydration shell was considered, which was determined in 3.5 Å between oxygen atoms of waters and heavy polar atoms of the interested protein region. 5 μs conventional MD trajectories were used for each monomer during the water analysis. The procedure for the MD simulations was described in the section "Conventional (unbiased) molecular dynamics simulations".

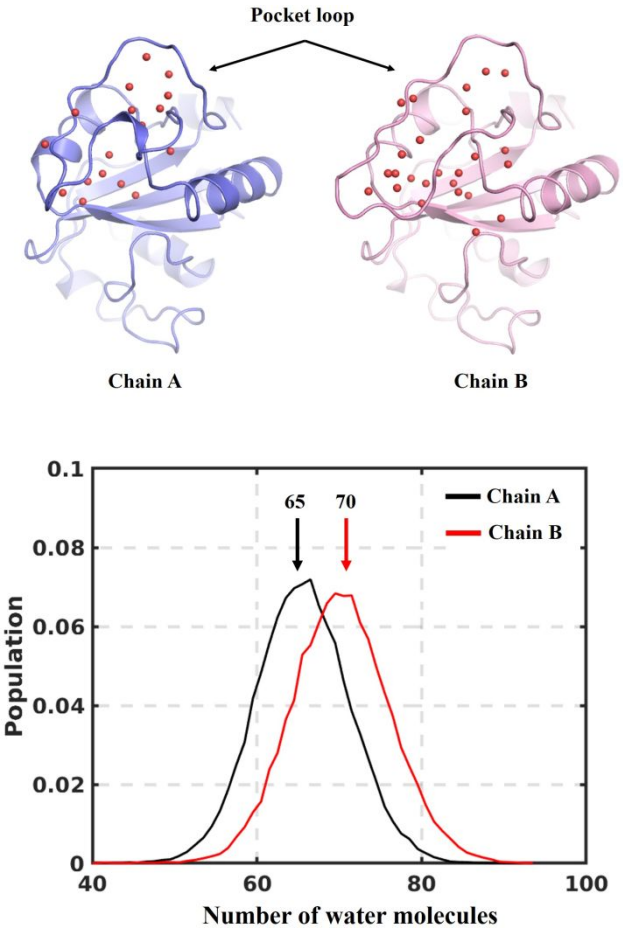


Figure S8. Root mean square fluctuation (RMSF) comparison between wild-type and Thr379Val mutant proteins in their apo forms. The average structure along 5 μs MD trajectory was used as the reference for calculating the RMSF plot for each system. To obtain the average structure, the crystal structure was first used as the reference structure, and all snapshots of trajectories were superposed to backbone atoms of the reference structure. During the superposition, only a relatively rigid part of the structure was included, that is, excluding N- and C- terminals and the pocket loop. Finally, heavy atoms of backbone were used for the RMSF calculation.

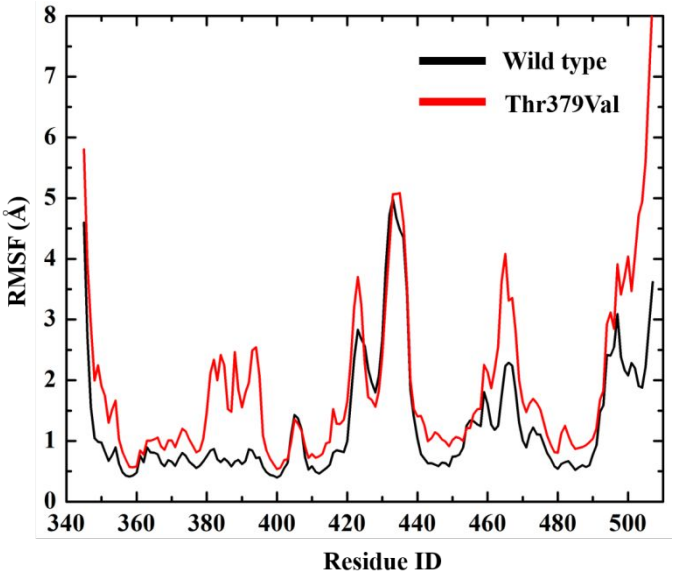
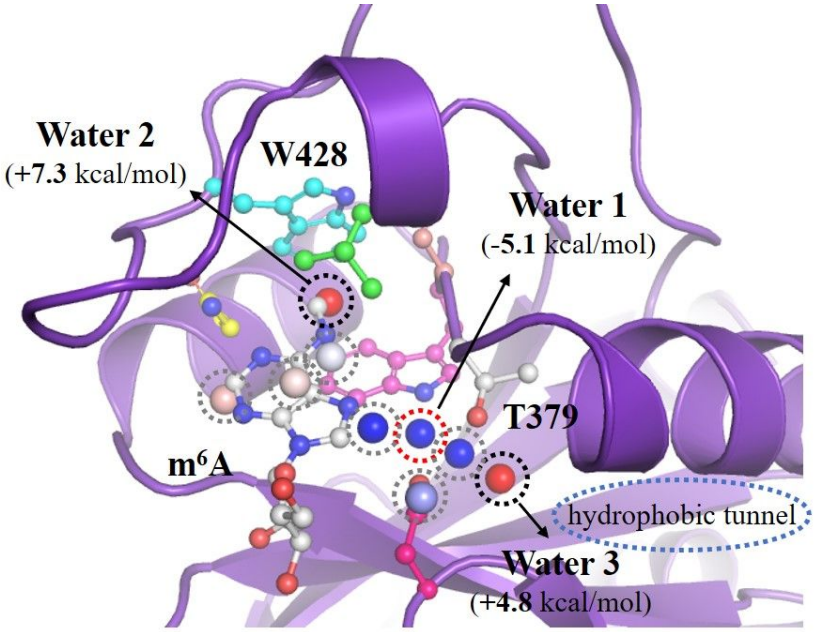
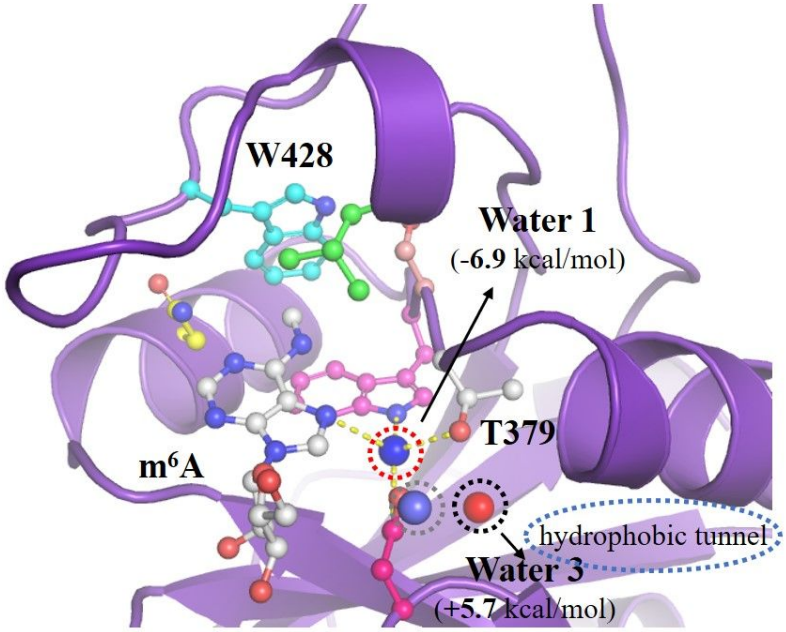


Figure S9. Water analysis of the m^{6}A binding site by the IST method. Predicted water molecules are shown (spheres with dashed circles). The calculated binding free energy of individual water molecules range from favourable (blue, difficult to displace by a ligand) to unfavourable (red, easy to displace). The m^{6}A structure is shown also in the apo form to illustrate the location of the water molecules displaced by the natural ligand.

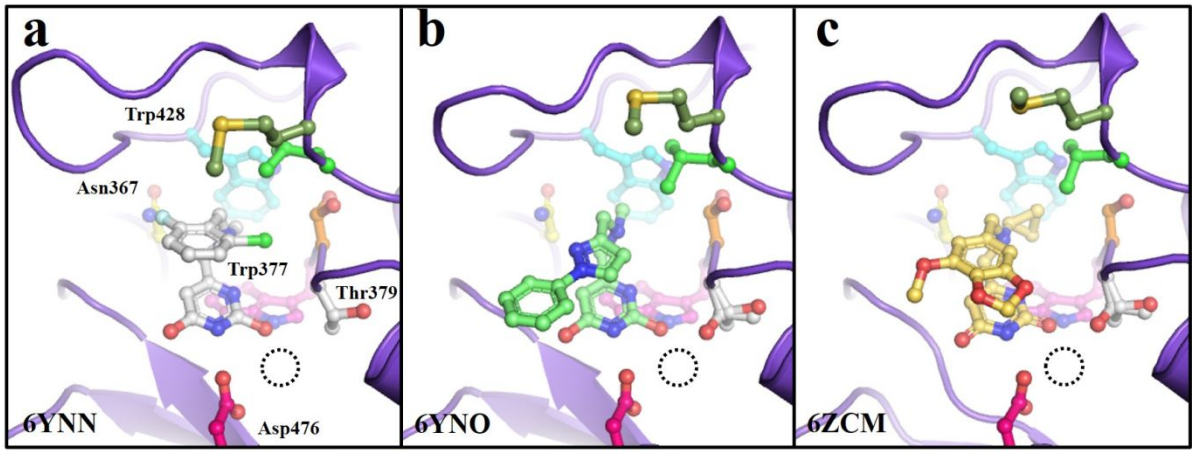


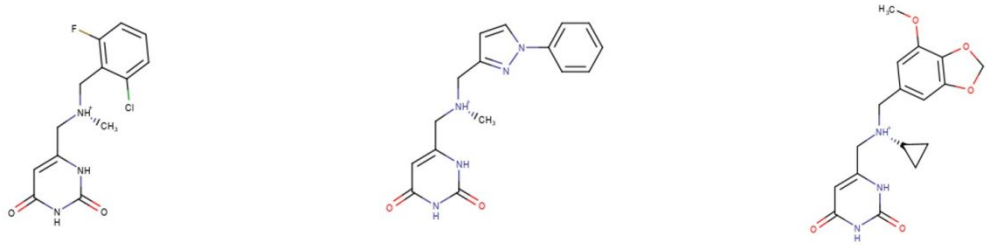
Apo



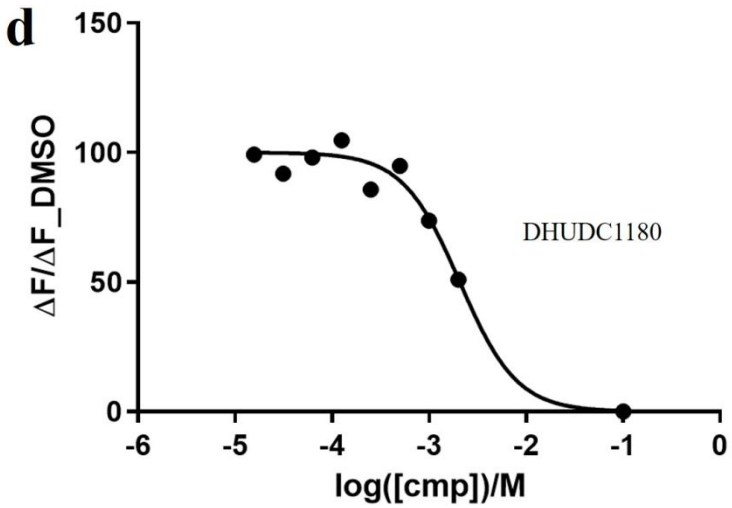
Holo

Figure S10. (a-c) Three small-molecule ligands of YTHDC1 that displace the key structural water molecule (water 1). The dashed circles indicate the position that water 1 has in the m^{6}A-YTHDC1 complex. These ligands have poor affinity as water 1 has optimal binding affinity according to IST calculations (see Figure S9). (d) The dose-response curve of the ligand DHUDC1180 can be fitted with an IC_{50} of 2 mM assuming 100% of signal reduction at 100 mM. The other two ligands do not show significant signal reduction at 2 mM.





DHUDC1135 DHUDC1139 DHUDC1180



Reference:

1. Xu, C.; Wang, X.; Liu, K.; Roundtree, I. A.; Tempel, W.; Li, Y.; Lu, Z.; He, C.; Min, J., Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **2014,** *10* (11), 927-9.
2. Kabsch, W., Xds. *Acta Crystallogr. D* **2010,** *66*, 125-132.
3. Mccoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40*, 658-674.
4. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **2012,** *68*, 352-367.
5. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D* **2010,** *66*, 486-501.
6. Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.* **1992,** *97* (4), 2635-2643.
7. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant-Pressure Molecular-Dynamics Algorithms. *J. Chem. Phys.* **1994,** *101* (5), 4177-4189.
8. Sun, X. Q.; Agren, H.; Tu, Y. Q., Functional Water Molecules in Rhodopsin Activation. *J. Phys. Chem. B* **2014,** *118* (37), 10863-10873.