Flexible binding of m⁶A reader protein YTHDC1 to its preferred RNA motif

Yaozong Li, Rajiv Kumar Bedi, Lars Wiedmer, Danzhi Huang, Paweł Śledź* and Amedeo Caflisch*

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

*To whom correspondence should be addressed. Tel: +41 44 635 5521; Fax: +41 44 635 6862; Email: caflisch@bioc.uzh.ch

Correspondence may also be addressed to Paweł Śledź. Tel: +41 44 63 55587; Email: p.sledz@bioc.uzh.ch

Yaozong Li's ORCID: 0000-0002-5796-2644

Rajiv Kumar Bedi's ORCID: 0000-0002-8193-9006

Paweł Śledź ORCID: 0000-0002-4440-3253

Amedeo Caflisch ORCID: 0000-0002-2317-6792

Abstract

N6-methyladenosine (m^6A) is the most prevalent chemical modification in human mRNAs. Its recognition by reader proteins enables many cellular functions, including splicing and translation of mRNAs. However, the binding mechanisms of m^6A -containing RNAs to their readers are still elusive due to the unclear roles of m^6A -flanking ribonucleotides. Here, we use a model system, YTHDC1 with its RNA motif 5'-G₋₂G₋₁(m^6A)C₊₁U₊₂-3', to investigate the binding mechanisms by atomistic simulations, X-ray crystallography, and isothermal titration calorimetry. The experimental data and simulation results show that m^6A is captured by an aromatic cage of YTHDC1 and the 3' terminus nucleotides are stabilized by cation- π - π interactions, while the 5' terminus remains flexible. Moreover, simulations of unbound RNA motifs reveal that the methyl group of m^6A and the 5' terminus shift the conformational preferences of the oligoribonucleotides to the bound-like conformation, thereby facilitating the association process. The proposed binding mechanisms may help in the discovery of chemical probes against m^6A reader proteins.

Introduction

Methylation of position N6 in adenines, also known as m⁶A modification, is the most prevalent internal modification of messenger RNA (mRNA) with approximately one in 200 adenine bases modified, and as such it makes significant contributions to central events in biology (Dominissini, Moshitch-Moshkovitz, Salmon-Divon, Amariglio, & Rechavi, 2013; Meyer et al., 2012). On the molecular level, the introduction of m⁶A affects structures of RNAs and their ability to form protein-RNA interactions, and thus modulates processing (Alarcon, Lee, Goodarzi, Halberg, & Tavazoie, 2015; Xiao et al., 2016), translation (Lin, Choe, Du, Triboulet, & Gregory, 2016; X. Wang et al., 2015), and stability of the cellular transcripts (X. Wang et al., 2014). As a consequence, m⁶A is implicated in controlling embryonic development processes and stem cell differentiation (Batista et al., 2014; Geula et al., 2015; Qi et al., 2016; Y. Wang et al., 2014), regulating the mammalian circadian clock (Fustin et al., 2013), and modulating stress response, e.g., heat shock (Zhou et al., 2015). The malfunctions of the cellular machinery regulating the m⁶A modification have in turn been linked to pathologies like obesity (Ben-Haim, Moshitch-Moshkovitz, & Rechavi, 2015), cancer (Peng, Yuan, Jiang, Tang, & Li, 2016), and neurodegeneration (Satterlee et al., 2014). Therefore, the m⁶A field has recently attracted broad attention of drug discovery community (Boriack-Sjodin, Ribich, & Copeland, 2018; Dai, Wang, Zhu, Jin, & Wang, 2018; Delaunay & Frye, 2019; Hsu, Shi, & He, 2017).

While the m⁶A modification has been known for several decades, only recently it has been proven reversible due to the discovery of the selective demethylases (erasers) that remove it from mRNA (Jia et al., 2011; Zheng et al., 2013) and balance the actions of the writer complex (J. Z. Liu et al., 2014; Sledz & Jinek, 2016; P. Wang, Duxtader, & Nam, 2016; X. Wang et al., 2017; Wiedmer, Eberle, Bedi, Sledz, & Caflisch, 2019). Similarly to the epigenetic modifications of histone proteins, the reversibility of the mRNA methylation gives it the potential to constitute an additional layer of epitranscriptomic regulation of gene expression (N. Liu & Pan, 2016; Meyer & Jaffrey, 2014; X. Zhang & Jia, 2016). The reader proteins, i.e., proteins that recognize the modification and elicit response upon it, play the key role at the executive end of such regulation mechanism. A number of proteins have been identified as direct m⁶A readers, with different structures, subcellular localization and mechanism of the recognition of epitranscriptomic mark. There are five YTH-domain-containing m⁶A readers that adopt highly similar open α/β fold and specifically recognize the m⁶A mark through interactions of aromatic cage (F. D. Li, Zhao, Wu, & Shi, 2014; C. Xu et al., 2015; C. Xu et al., 2014; Zhu et al., 2014). Cytoplasmic readers YTHDF1 and YTHDF3 play roles in translation initiation (X. Wang et al., 2015), while highly similar YTHDF2 reader controls destabilization of cellular transcripts by targeting them to P-bodies (X. Wang et al., 2014). The nuclear reader YTHDC1 plays a role in splicing regulation (Xiao et al., 2016), while YTHDC2 contains additional RNase helicase domain and is important in meiosis (Wojtas et al., 2017).

Specific recognition of m⁶A-containing transcripts by the YTH-domain readers has been studied in details by the means of protein crystallography, pointing out at rigid and conserved recognition of consensus sequence GG(m⁶A)CU. Beyond YTH-domain m⁶A readers, there are also indirect readers (i.e., proteins not interacting with m⁶A but binding to m⁶A-containing transcripts) including HBRNPC (N. Liu et al., 2015), and anti-readers like G3BP1 that bind preferably to unmodified transcripts (Vermeulen, 2017). The crucial role of m⁶A in the molecular recognition has been confirmed, where an aromatic cage consisting of number of key tryptophan residues enables binding. However, the exact role of m⁶A-flanking nucleotides that make a number of hydrogen bonds and salt bridges to the protein (S. K. Luo & Tong, 2014; C. Xu et al., 2014), remains elusive. Further studies on these interactions are crucial not only for understanding the mechanism and timescale of epitranscriptomic regulation, but also for informing the development of therapeutic agents.

It is not easy to infer the importance of the interactions between the m⁶A-flanking nucleotides and reader proteins, partially because they are observed on protein surface and thus may be stabilized by crystal packing. In addition, unbound oligoribonucleotides are highly flexible and may adopt many metastable conformations in aqueous solution, which is difficult to study by experiments. To this end, molecular dynamics (MD) may provide complementary information for understanding the binding mechanisms (Dror, Dirks, Grossman, Xu, & Shaw, 2012; Karplus & McCammon, 2002; Sledz & Caflisch, 2018). However, the field of MD simulations on RNA and its complexes with proteins, e.g., m⁶A reader proteins and its ligand RNAs, is less mature than those on proteins (or with organic molecules). In

particular, three potential issues hinder the use of RNA-related force fields (Krepl et al., 2015; MacKerell & Nilsson, 2008; Sponer et al., 2017; Vangaveti, Ranganathan, & Chen, 2017): I: the inaccuracy of ribonucleotide force fields themselves (Krepl et al., 2015; Sponer et al., 2017), II: lack of parameters for modified nucleotides (Aduri et al., 2007; Y. Xu, Vanommeslaeghe, Aleksandrov, MacKerell, & Nilsson, 2016) and III: the difficulties in balancing interactions between proteins and RNA molecules (Krepl et al., 2015; MacKerell & Nilsson, 2008).

The first issue has been moderately alleviated by the recent improvements of common RNA force fields, e.g., CHARMM (Denning, Priyakumar, Nilsson, & MacKerell, 2011), AMBER (Tan, Piana, Dirks, & Shaw, 2018) and OPLS (Robertson, Qian, Robinson, Tirado-Rives, & Jorgensen, 2019). For this purpose, different methodologies have been explored, for example refining bonded and non-bonded potentials (Denning et al., 2011; Robertson et al., 2019; Tan et al., 2018) and adding external hydrogen terms (Kuhrova et al., 2019). The improved force fields have led to successful implementations of MD simulations on some RNA-containing systems, including conformational dynamics of folded RNAs (Banas et al., 2010; Krepl, Clery, Blatter, Allain, & Sponer, 2016; Kuhrova et al., 2016; Musiani et al., 2014; Salmon, Bascom, Andricioaei, & Al-Hashimi, 2013; Templeton & Eber, 2018; Y. Xu, MacKerell, & Nilsson, 2016) and biological catalysis (Banas, Jurecka, Walter, Sponer, & Otyepka, 2009; K. Nam, Gao, & York, 2008; K. H. Nam, Gaot, & York, 2008). Concerning the second issue, naturally occurring modified ribonucleotides parameters have now been developed and are consistent with commonly used force fields (Aduri et al., 2007; Y. Xu, Vanommeslaeghe, et al., 2016). The last issue may be most critical for the successful simulation of a protein-RNA complex. For balancing protein-RNA interactions, one strategy is to modify pair-specific corrections to non-bonded interactions by targeting related osmotic pressure (Y. Luo & Roux, 2010; J. Yoo & Aksimentiev, 2016, 2018; J. J. Yoo & Aksimentiev, 2012). Besides, polarizable force fields may capture essential physical effects for protein-RNA interactions, but their developments are not fully mature yet (Lemkul & MacKerell, 2018; Yuan, Zhang, Mills, Hu, & Zhang, 2018; C. S. Zhang et al., 2018).

Here, we use YTHDC1 protein as a model system to investigate the binding of the preferred RNA motif to m⁶A-reader proteins by protein crystallography, isothermal titration calorimetry (ITC), and explicit solvent MD simulations with a force field suitable for protein-RNA complexes. The role of upstream and downstream m⁶A-flanking nucleotides in the binding event is analyzed at the atomistic level. The MD simulations of the bound states reveal the flexibility of the 5' terminal nucleotides, and the role in stabilizing cation- π - π interaction by the 3' terminal nucleotides. From the unbound-state simulations, we found that the methyl group of m⁶A and 5' terminus shift the nucleotides' conformations into the bound-like one observed in the structures of the complex. Taken together, the crystal structures and simulation results of bound and free oligoribonucleotides provide an atomistic explanation for the roles of m⁶A-flanking ribonucleotides in the molecular recognition.

Materials and methods

Protein purification

The plasmid expressing the N-terminally hexahistidine-tagged YTH domain (residues 345-509) of human YTHDC1 protein was obtained as a gift from Cheryl Arrowsmith (Addgene ID: 64652). The recombinant protein was purified to homogeneity in two chromatographic steps. The protein was overexpressed for 16 hours at 24°C in *Escherichia coli* BL21 (DE3) cells upon induction with 0.4 mM IPTG. The cells were harvested and resuspended in the lysis buffer containing 100 mM Tris-HCl at pH 8.0, 500 mM NaCl and 10 mM imidazole. The cells were lysed by sonication and the cell lysate was clarified by centrifugation at 48,000 g for one hour and loaded onto Ni-NTA affinity column (5 mL HisTrap FF from GE Healthcare). After extensive washing with the wash buffer containing 100 mM Tris-HCl at pH 8.0, 500 mM NaCl and 50 mM imidazole the target protein was eluted with elution buffer containing 100 mM Tris-HCl at pH 8.0, 500 mM NaCl and 250 mM imidazole. The N-terminal hexahistidinetag was removed by cleavage with tobacco etch virus (TEV) protease at 1:50 ratio. The excess imidazole was removed by overnight dialysis and the sample was subjected to secondary subtractive Ni-NTA affinity chromatography step to remove the protease and uncleaved protein. Finally, the protein was subjected to a gel filtration step using Superdex 75 16/60 column in a buffer containing 10 mM Tris-HCl at pH 7.5, 150 mM NaCl and 1 mM DTT. The protein was concentrated to 10 mg/mL, flash-frozen in liquid nitrogen and stored at -80 °C for future experiments.

Crystallography

The crystals of YTHDC1 YTH domain 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 oligoribonucleotides (custom-synthesized and HPLC-purified by Dharmacon), the crystals were transferred to a 1 µL drop containing 10 mM oligoribonucleotide 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 (Kabsch, 2010). The structures were solved by molecular replacement using Phaser program (McCoy et al., 2007) from the Phenix package (Afonine et al., 2012). The unliganded structure of YTHDC1 (PDB ID: 4R3H) was used as a search model. The model building and refinements were performed using COOT (Emsley, Lohkamp, Scott, & Cowtan, 2010) and phenix.refine (Afonine et al., 2012). Data collection and refinement statistics are summarized in Table S1.

Coordinates

The atomic coordinates and structure factors for the four determined complexes have been deposited in the Protein Data Bank (PDB) with the accession codes 6RT4, 6RT5, 6RT6 and 6RT7.

Binding assays

Isothermal titration calorimetry (ITC) experiments were carried out at 18°C using MicroCal ITC200 (GE Healthcare). For the ITC experiments, the final gel filtration step of the purification of YTH domain of YTHDC1 was performed in the ITC buffer (20 mM HEPES, pH 7.5, 150 mM NaCl). Oligoribonucleotides were directly dissolved at 500-800 µM in the ITC buffer and titrated into the sample cell containing the protein at 50-60 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 analyzed using a single binding site model, provided in the MicroCal Origin software package.

Molecular dynamics simulations

All the systems for MD simulations in this study were constructed based on the structure of the YTH domain of YTHDC1 complexed with modified pentaribonucleotide reported previously by others (PDB ID: 4R3I) (C. Xu et al., 2014). The protein YTHDC1, the RNA oligomer GG(m⁶A)CU, and crystal water

molecules were kept as in the initial structure. Based on this structure, six different systems were generated, i.e., three complex systems and three structures of unbound modified oligoribonucleotides. For the three complex systems, the protein coordinates were kept unchanged, while the oligoribonucleotides were truncated to obtain the tetrameric G(m⁶A)CU and GG(m⁶A)C from the pentameric GG(m⁶A)CU, whose coordinates were available in the crystal structure. The hydrogen atoms were added by the CHARMM program (Brooks et al., 2009) and the protonation states were determined at neutral pH conditions. The complex systems were solvated in a 67 Å of rhombic dodecahedron (RHDO) TIP3P water box (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) to ensure 10 Å buffer space between the macromolecular atoms and the boundary of the water box. To neutralize the system and mimic the physiological conditions, Na⁺ and Cl⁻ ions at 0.15 M concentration were added to the solvated systems. Similarly, the three free RNA ligands were solvated in a 47 Å of RHDO water box with 0.15 M of NaCl. To explore the regulatory role of the m⁶A's methyl group in RNA's conformations, the unmethylated pentamer GGACU was also simulated in its free state.

Each simulation system was initially minimized for 10,000 steps under a series of restraints and constraints on the solute atoms to release its bad contacts and poor geometry. The minimized structure was heated to 300 K and equilibrated in NVT condition (constant volume and temperature). Finally, the structure was further equilibrated in NPT condition (constant pressure and temperature). All the equilibration phases lasted for 1 nanosecond (ns) using the CHARMM program (version 42b2) (Brooks et al., 2009). Production runs of 500 ns each were carried out in NPT conditions using the NAMD program (version 2.12) (Phillips et al., 2005). The pressure was controlled by Nosé-Hoover Langevin piston method with 200 picosecond (ps) piston period and 100 ps piston decay time (Feller, Zhang, Pastor, & Brooks, 1995; Martyna, Tobias, & Klein, 1994). The temperature was maintained at 300 K using the Langevin thermostat with a 5 ps friction coefficient. The integration time step was set to 2 fs by constraining all the bonds involving hydrogen atoms by the SHAKE algorithm. Van der Waals energies were calculated using a switching function with a switching distance from 9 to 11 Å (Steinbach & Brooks, 1994), and electrostatic interactions were evaluated using the particle mesh Ewald summation (PME) method (Essmann et al., 1995).

The CHARMM36 force field was used for protein (Huang & MacKerell, 2013) and RNA (Denning et al., 2011) molecules. For the parameterization of m⁶A, the force field for naturally occurring modified ribonucleotides was used (Y. Xu, Vanommeslaeghe, et al., 2016). Five independent runs with random initial velocities were carried out for each system for a total of 2.5 microseconds (μs). MD snapshots were saved every 20 ps along the MD trajectories for further analysis. Geometric measurements, e.g., distance, dihedrals, root mean square deviation (RMSD) and contact map analysis, were performed with CHARMM routines. All statistical figures were plotted by MATLAB (Version 2018a MathWorks, Inc.) and structural figures were generated with the PyMOL graphic software (Version 2.2 Schrödinger, LLC.).

Results

Flexible binding of the G₂G₋₁ segment of the GG(m⁶A)CU ligand to the YTH domain

The crystal structure of the complex between the pentamer GG(m⁶A)CU and YTHDC1 YTH domain (PDB ID: 4R3I) shows that m⁶A is fully buried in an aromatic cage consisting of Trp377 and Trp428 (Figure 1) (C. Xu et al., 2014). On the other hand, the contribution to binding of the non-modified nucleotides is less clear. The two guanines of GG(m⁶A)CU are visible in the structure 4R3I, interacting with the surface residues of YTHDC1 (Figure 1A). The guanine bases of G-1 and G-2 form hydrogen bonds (HB) with the backbone of Val382 and Asp476, respectively (Figure 2A). Furthermore, there is a HB between the two guanines (C. Xu et al., 2014). To test the stability of these interactions, we conducted MD simulations on the complex system of YTHDC1 with GG(m⁶A)CU. The interaction analysis shows that the HB interactions are not well maintained during the simulations with large deviation from the distances in the crystal structure (Figure S1).

Figure 1. Binding mode of the RNA motif GG(m⁶A)CU to the YTHDC1 reader domain (PDB ID: 4R3I). (A) The backbones of RNA (yellow) and YTHDC1 (red, blue, and green for α -helices, β -strands, and loops, respectively) are shown in cartoon and the two tryptophanes of the aromatic cage are shown in sticks. The binding pocket of m⁶A is highlighted (black dashed circle). (B) Detailed interactions between m⁶A (carbon atoms in gray) and the pocket of YTHDC1 (tryptophane and polar residues in salmon and green, respectively). The importance of the two tryptophanes for the binding has been confirmed by site-directed mutagenesis experiments (C. Xu et al., 2014).

We checked whether crystal packing plays a role in stabilizing the HB interactions between the nucleotides G₋₁G₋₂ and YTHDC1. In particular, several interactions are found at the interface between molecules related by crystallographic symmetry (Figure 2B). For example, towards the RNA's 5' end, the sugar group of G₋₂ interacts with the side chain of Glu405 of the monomer related by translation and the 2₁ screw axis. Similarly, the phosphate linker of G₋₁ in the gray monomer interacts with the sugar group of U₊₂ in the blue monomer and is further stabilized by a salt bridge contributed by Lys472. Such contacts of oligoribonucleotides from two YTHDC1-RNA complexes are unlikely populated in aqueous environment, and therefore they are probably a consequence of crystal packing. Similar packing modes are also found in other structures of reader proteins (Figure S2).

Figure 2. Interactions stabilized by crystal packing. **(A)** Hydrogen bonds between nucleotides (G₋₁ and G₋₂) and YTHDC1. The interaction mode was extracted from the structure 4R3I. The plausible hydrogen bonds are indicated in red dotted lines. **(B)** Two m⁶A-containing RNA oligomer-YTHDC1 complexes related by crystallographic symmetry, shown and labeled in gray and blue for clarity. Their RNA structures are presented in sticks. The protein residues that contribute the crystal packing interactions are colored in yellow. **(C)** New complex structures solved in the present study. The proteins of the complex structures are shown in gray cartoon and the bound RNAs are shown in sticks with different colors. The residues in the m⁶A binding site are shown in sticks (same colors as in Figure 1B).

We hypothesized that truncated variants of GG(m⁶A)CU, i.e., without terminal nucleotides, would not permit the crystal contacts mentioned above. To test this, four different RNAs were synthesized for crystallographic study, i.e., GG(m⁶A)C, G(m⁶A)CU, G(m⁶A)C, and (m⁶A)CU. We solved the crystal structures of the YTH domain of YTHDC1 in the complex with these four RNA oligonucleotides at resolutions of 1.5, 1.7, 2.3 and 1.5 Å, respectively (Table S1). The new structures mostly agree well with the structure of the complex with the pentamer RNA ligand (Figure 2C). However, the coordinates of the two guanine nucleotides, including bases, sugars and phosphates, at positions -2 and -1 with respect to m⁶A cannot be modelled in our structures due to missing electron density. By contrast, the cytosine and uridine at positions +1 and +2, respectively, show well defined electron density. Thus, our co-crystal structures suggest that the two guanine ribonucleotides upstream of m⁶A do not form stable interactions with the protein, while the downstream motif CU forms conserved interactions with the protein. These findings are consistent with our hypothesis on the stabilization of the untruncated GG(m⁶A)CU due to crystal packing in the YTHDC1 structure (PDB id: 4R3I).

5' and 3' terminal nucleotides contribute to the affinity of GG(m⁶A)CU to YTHDC1

The four crystal structures have provided evidence that the G₋₂G₋₁ dinucleotide segment does not form stable interactions with YTHDC1 while the segment C₊₁U₊₂ shows conserved contacts with the protein. However, it is not clear if and how much these terminal nucleotides contribute to the binding of GG(m⁶A)CU to YTHDC1. To address this question, we measured dissociation constant (*K_d*) values of three representative oligomers, viz., GG(m⁶A)CU, G(m⁶A)CU, and GG(m⁶A)C using isothermal titration calorimetry (ITC) (Figure 3 and Table S1). We obtained a *K_d* value of 0.5 μM for GG(m⁶A)CU, which is similar to a value reported previously (2.0 μM) (C. Xu et al., 2014). The U₊₂-deficient tetramer, i.e., GG(m⁶A)C, is less potent by a factor of about 20 than the pentamer GG(m⁶A)CU. Such binding

reduction can be intuitively understood from the crystal structures, namely U_{+2} forms π - π stacking with C_{+1} and thus stabilizes interactions of the 3' end with YTHDC1. Surprisingly, the G_{-2} -deficient oligomer, i.e., $G(m^6A)CU$, shows significant reduction of affinity (~ 7 times compared to the pentamer) although G_{-2} does not form stable interactions with YTHDC1 in our crystal structures and MD simulations (Figure 1C and S1). Thus, further investigation is needed to understand how G_{-2} impacts the binding affinity (See below).

Figure 3. Isothermal titration calorimetry measurements of the binding of the RNA oligomers $GG(m^6A)CU$, $G(m^6A)CU$, and $GG(m^6A)C$ to YTHDC1. The raw ITC curves and binding isotherms with the best fit are shown in the top and bottom panels, respectively. The truncation of one nucleotide at either 5'- or 3'-terminus of $GG(m^6A)CU$ deteriorates the affinity.

MD simulations of liganded YTHDC1 reveal flexibility of the nucleotides upstream of m^6A

Three complexes with ligand $GG(m^6A)CU$, $G(m^6A)CU$, and $GG(m^6A)C$ were investigated by MD simulations to analyze their relative flexibility. In all simulations, the two guanine nucleotides upstream of m^6A show large deviations compared to their crystal poses (Figure 4 and Figure S3). This result is consistent with the lack of electron density for this RNA segment. During the MD simulations, the two guanine nucleotides do not show stable and specific interactions with YTHDC1 in contrast with the crystal structure (PDB code 4R3I) (Figure 2A). Instead, the two nucleotides are dynamically coupled for a long time by the π - π interaction between their bases. This is illustrated by the contact map analysis, in which G_{-1} and G_{-2} interact with each other in approximately half of the samplings (lower triangles of Figure 5).

The tetranucleotide ligands allow us to assess the role of the terminal G_{-2} and U_{+2} , without which m^6A -containing RNAs show reduced binding affinity compared to the pentamer. Although the nucleotide G_{-2} was flexible during the simulations of the complex, it helps stabilizing the pose of m^6A in the binding pocket of YTHDC1. Without the terminal guanine nucleotide at position -2, m^6A experiences slightly larger deviations and fluctuations than in G_{-2} -containing ligands (Figure 4 and Figure S4). The larger RMSD values are due to the partial dissociation of m^6A from the YTHDC1 binding pocket in two of the five trajectories of $G(m^6A)CU$, which can be monitored by three key hydrogen bonds between m^6A and YTHDC1 (Figure S4). Noteworthy, if the dissociated snapshots are neglected in the system $G(m^6A)CU$, the RMSD value of m^6A is nearly unchanged compared to other two systems (Figure S5). Therefore, the reduced affinity of the G_{-2} deficient ligand may not be plainly explained based on simulation results of the bound states. Other factors causing the affinity reduction will be investigated in the later sections.

Figure 4. RMSD profile of RNA oligomers in the bound state. The plots show average RMSD value and standard deviation calculated over the cumulative sampling of five trajectories for each simulation system. To calculate the RMSD value, heavy atoms of nucleotides in their crystal structure (PDB ID: 4R3I) were used as reference. All snapshots from MD simulations were superimposed to the protein backbone of the crystal structure excluding C- and N-terminal segments.

Concerning the nucleotides downstream of m⁶A, our crystal structures and MD simulations provide evidence that C₊₁ and U₊₂ form more stable interactions with YTHDC1 than G₋₁ and G₋₂. The downstream C₊₁ establishes cation- π and π - π interactions with Arg475 and U₊₂, respectively. A previous study has confirmed the importance of the cation- π interaction for the molecular recognition between YTHDC1 and GG(m⁶A)CU by mutating Arg475 to phenylalanine or alanine with 10-fold and 100-fold reduction, respectively, compared to the wild type (C. Xu et al., 2014). However, the role of U₊₂ has not been confirmed as it is stabilized by the crystal packing in the structure 4R3I (Figure 2B). The simulations for the tetrameric ligand GG(m⁶A)C show that the C₊₁-Arg475 interaction becomes less stable than in the case of the ligand with U₊₂. Consistently, the RMSD of C₊₁ becomes much larger and fluctuates more than that of the U₊₂-containing oligomers (blue bars in Figure 4). This observation indicates that U₊₂ indeed stabilizes the interaction between C₊₁ and Arg475. As shown in Figure S6, the contact distance for C₊₁ and Arg475 is locally distributed at 3.8 Å for U₊₂-containing ligands (black and red lines in the Figure S6), while it becomes loosely distributed in the U₊₂-deficient ligand (the blue line). These structural and dynamic analyses rationalize the role of U₊₂ and the reduced binding of the tetrameric ligand GG(m⁶A)C (shown by ITC data).

The C₊₁U₊₂ segment specifically interacts with a positively charged shallow pocket on the protein surface and may play a comparably important role in the protein-RNA recognition compared to m⁶A. An equivalent mutation to Arg475Ala was also reported for a homologous protein YTHDF2, i.e., Arg527Ala. This mutation showed approximately 25-fold reduction of the binding affinity for its RNA oligoribonucleotides compared to the wild type (Zhu et al., 2014). Moreover, the C₊₁U₊₂ segment binds to a well-defined pocket on the YTHDC1 surface, consisting of four positively charged residues, i.e., Lys361, Arg404, Lys472, and Arg475 (Figure 1A). The binding data and structural analysis suggest that the interaction between the C₊₁U₊₂ segment and the reader protein is specific rather than an undirected charge-charge interaction. In addition, Zhu *et al.* suggested that a two-step model fits the data from a fluorescence polarization assay better than the one-step model for the binding of YTHDF2 to an m⁶A-containing oligoribonucleotide (Zhu et al., 2014). They proposed that the binding of m⁶A-flanking nucleotides facilitates the recognition of m⁶A. Integrating this two-step model with our results from ITC and MD simulations, we infer that the C₊₁U₊₂ segment plays an essential role in the association process of GG(m⁶A)CU to DC1 comparable to that of m⁶A. However, it is still not clear which the first motif is to be recognized by the reader protein, m⁶A or C₊₁U₊₂ segment, and this question requires further studies.

Conformational heterogeneity of m⁶A-containing RNAs in bound and free states

To rationalize relative binding affinities, we analyzed conformations of three RNA oligomers in their bound and free states. Significant differences are found in the intramolecular contact pattern between

the bound and free states (Figure 5). The bound RNAs show reduced contacts between nucleotides compared to their free states. For example, three pairs of nucleotides in GG(m⁶A)CU's free state are in contact for more than 40% of all snapshots. Its bound state, by contrast, loses one stable interaction pair, i.e., G₋₁-C₊₁ interaction pair (Figure 5). That is because the centrally located m⁶A of the bound state firmly anchors in the aromatic cage of YTHDC1. Thus, the pentameric GG(m⁶A)CU is divided into two independent segments, i.e., G₋₁-G₋₂ and C₊₁-U₊₂, separated by m⁶A. Two nucleotides belonging to the same segment are frequently coupled, but nucleotides from different segments interact much less (lower triangles in Figure 5).

Figure 5. Contact maps of m⁶A-containing RNA oligomers in their free state (upper, left triangle) and in the complex with YTHDC1 (lower, right triangle). Pairwise contact frequencies are evaluated by the distance of center of mass between two bases and a cutoff value of 5 Å. The color bar represents the frequency of contact between two nucleotides, ranging from dark blue (no interaction) to dark red (persistent interaction). The nucleotide pairs which are not included in the systems are presented in gray squares.

The purine base of m⁶A fully interacts with the aromatic cage of YTHDC1. Therefore, it is reasonable to expect that m⁶A should be sufficiently exposed to the solvent before the association with the protein, so that its binding pocket can effectively recognize m⁶A. The contact maps show that m⁶A in the G₋₂-containing ligands, i.e., GG(m⁶A)CU and GG(m⁶A)C, is not in frequent contact with other nucleotides in the unbound state. By contrast, m⁶A in G(m⁶A)CU does interact with C₊₁ (upper triangles in Figure 5). The distribution of m⁶A-C₊₁ distance for G(m⁶A)CU is more populated for values around 5 Å than for the other two ligands (Figure S7).

To detect how the exposure of m⁶A varies in different ligands, we measured the solvent accessible surface area (SASA) of m⁶A in the free state (Figure S8A). The two G₋₂-containing ligands show a similar distribution of SASA, which is mostly populated around 180 Å² with a second peak at 200 Å². By contrast, G(m⁶A)CU populates its main peak at 150 Å². Such difference in SASA is caused by the frequent contact between m⁶A and C₊₁ (the upper triangle for G(m⁶A)CU in Figure 5). Taken together, the G₋₂ nucleotide plays a role in the exposure of m⁶A from the RNAs in the unbound state, and thereby benefits the recognition of m⁶A by its reader proteins. To study if the methyl group of m⁶A facilitates the base exposure, we also compared the oligomers GG(m⁶A)CU and GGACU in their unbound states. Again, a significant difference was found in the distribution of the SASA, with the unmethylated oligomer showing a less exposed adenine than its N6-methylated counterpart (Figure S8B).

Conformational free-energy profiles of m⁶A-containing RNAs in the unbound state

Multiple MD simulations of the unbound state of GG(m⁶A)CU, G(m⁶A)CU, and GG(m⁶A)C were carried out to analyze potential differences in the most populated conformers. The simulations were used to construct free-energy profiles with the RMSD of (m⁶A)C relative to its crystal pose (Changeux & Edelstein, 2011) as the progress variable. The free-energy profiles show that oligomers in aqueous solution adopt multiple meta-stable conformations, which appear as local minima of the free-energy profiles (Figure 6). The ligands GG(m⁶A)CU and GG(m⁶A)C show a similar conformational propensity. There is approximately a 1 kcal/mol free-energy difference between their global minimum and the bound-like local minimum (black and blue lines in Figure 6). By contrast, G(m⁶A)CU experiences a ~2

kcal/mol of energy penalty from the unbound to the bound state. The conformational free-energy difference indicates that the former two ligands have less energy penalty for redistributing their conformations to the bound-state than G(m⁶A)CU. It is useful to introduce a two-state approximation in which all conformations of an RNA oligomer are divided into either bound or unbound states. The two-state model shows a significant energy difference between G₋₂-containing oligomers and G(m⁶A)CU irrespective of the RMSD threshold (Figure S9). A similar difference was also found by comparing the PMF profiles of unbound GG(m⁶A)CU and GGACU (Figure S8C), indicating that the methyl group of m⁶A helps in adopting the bound conformation.

Figure 6. Conformational landscape of m⁶A-containing RNA oligomers in the unbound state. (A) Potential of mean force (PMF) along the RMSD from the bound state. The structure of the RNA segment present in the three simulated oligonucleotides, i.e., (m⁶A)C, in the complex with YTHDC1 (PDB code 4R3I) is selected as reference for the bound state. The PMF is calculated as $-k_B T \ln \frac{p_i}{p_0}$, where k_B is Boltzmann factor, T is temperature (300 K), p_i is the population in bin i , and p_0 is the probability of the most populated bin (with a 0.1 Å of bin width). The labels M, I, and B indicate mostly populated, intermediate, and bound-like conformational states, respectively. (B) Representative structures of different conformational states of free G(m⁶A)CU (top) and GG(m⁶A)CU (bottom).

The analysis on meta-stable conformations suggests an important role of G₋₂ in maintaining the bound-like conformations and exposing m⁶A to the bulk water. For example, for the oligomer GG(m⁶A)CU, the global minimum conformation along the free-energy profile not only maintains a bound-like pose but also exposes m⁶A to the solvent (M_{GG(m6A)CU} in Figure 6B). In this conformation, G₋₂ interacts with G₋₁ in a π - π interaction and thereby frees m⁶A to expose to the bulk water. In the most stable conformation of G(m⁶A)CU, G₋₁, m⁶A and C₊₁ stack together and thus m⁶A is less exposed to the solvent (the upper panel in Figure 6B). Although the conformational change from M_{G(m6A)CU} to I_{G(m6A)CU} increases the solvent exposure of m⁶A, C₊₁ and G₋₁ become packed with each other, resulting in a detrimental pose for its binding to YTHDC1. Finally, to transform this intermediate state to the bound-like one, the ligand G(m⁶A)CU has to pay a higher free-energy cost than GG(m⁶A)CU. Thus, G₋₂ is a key nucleotide for modulating the conformational preference of m⁶A-containing RNAs in the unbound state.

Discussion

We aimed to elucidate the roles of m⁶A-flanking nucleotides in the molecular recognition of YTHDC1 to its RNA motif GG(m⁶A)CU by atomistic simulations, X-ray crystallography, and isothermal titration calorimetry. The selected model oligoribonucleotide is biologically relevant because the sequence is preferentially recognized by all five human m⁶A reader proteins (A. Li et al., 2017; F. D. Li et al., 2014; Shi et al., 2017; C. Xu et al., 2014). To investigate the flexibility of terminal nucleotides, the structures of the YTH domain of YTHDC1 in the complex with GG(m⁶A)C, G(m⁶A)CU, G(m⁶A)C, and (m⁶A)CU were solved at high resolution. ITC binding experiments were utilized to examine the contribution of 5' and 3' terminal nucleotides to the affinity of GG(m⁶A)CU for YTHDC1. MD simulations were carried out for bound and unbound states of three representative RNA ligands, i.e., GG(m⁶A)CU, G(m⁶A)CU, and GG(m⁶A)C. The experimental and simulation results complement each other and provide a consistent and detailed picture of the binding mechanisms.

Crystal packing artificially stabilizes the poses of m⁶A-flanking oligoribonucleotides bounded to reader proteins

The contacts due to crystal packing play a non-negligible role in stabilizing the binding pose of m⁶A-flanking nucleotides with YTHDC1. As shown in Figure 2B, two 5' terminal nucleotides (G₂G₋₁) in an asymmetric unit interplay with another unit by multiple artificial contacts due to crystal packing, including hydrogen bonds and salt bridges. Although some interactions are also found between YTHDC1 and the m⁶A-flanking nucleotide in the same unit (Figure 2A), they may not be strong enough to maintain the observed crystal binding pose. The crystal packing could be still a critical stabilizing factor for the crystal pose. For testing this hypothesis, we soaked varied m⁶A-containing oligoribonucleotides into apo crystals of YTHDC1. Its space group does not permit artificial intramolecular contacts of oligoribonucleotides from different asymmetric units but still leave enough space for their natural interactions with the protein in the same unit. As a result, the G₂G₋₁ segment is disordered in all soaking structures, indicating weak interactions between them and the protein surface. The MD simulations further support this observation as the G₂G₋₁ segment significantly deviates from the crystal pose and shows substantial flexibility (Figure 4 and S1). Our evidence suggests a dynamical model for the G₂G₋₁ segment binding to YTHDC1 against the static model which seems to emerge from the previously published structure (4R3I)(C. Xu et al., 2014).

The available structural data suggest that the binding poses of 5' terminal nucleotides are often stabilized by crystal packing, and thus their crystal poses, and corresponding interactions should be carefully interpreted. Structural artifacts due to crystal packing are observed not only for YTHDC1, but also for other reader domains, leading to well-defined binding poses of m⁶A-flanking nucleosides, which otherwise are flexible in solution. We found in the PDB Data Bank another two m⁶A reader protein structures that are complexed with m⁶A-containing nucleotides, and they show multiple packing contacts between their m⁶A-flanking nucleotides and the protein surface (Figure S2). For example, the G₂G₋₁ segment in the structure 4RCJ (YTHDF1) interacts with the protein residue Ser461 and Ala462 of another asymmetric unit by multiple hydrogen bonds. Moreover, in the structure of YTH domain of MRB1 protein from *Z. roux* (PDB code 4U8T), the m⁶A-flanking nucleotides from different protein copies are even tangled, resulting in substantial artificial contacts (S. K. Luo & Tong, 2014). In addition, in an in-house YTHDF3 structure (to be published), the G₂G₋₁ segment is also stabilized by its own copies from another asymmetric unit.

Conformational analysis of oligoribonucleotides in their unbound state is essential for understanding the recognition mechanism

Our study reveals that the 5' terminus of GG(m⁶A)CU contributes to the affinity for YTHDC1 mainly because of the role of the 5' terminus in the unbound state rather than its direct interactions with the protein in the bound state. The importance of the terminal nucleotide G₂ for the affinity is supported by our ITC measurements, which show a 7-fold reduction of the G(m⁶A)CU motif with respect to the pentameric GG(m⁶A)CU (Figure 3 and Table S2). As mentioned above, the attenuated binding affinity is not due to the missing interactions between the 5' terminus and YTHDC1's protein surface. It is rationalized by MD simulations of different oligoribonucleotides in aqueous solution. The corresponding free energy landscape analysis demonstrates that the 5' terminal nucleotide G₂ contributes to the

binding by facilitating the adoption of bound-like conformations of the RNA oligomers (Figure 6), thereby reducing the free energy expense for the association process. These results show that information from both complex structures and conformations of oligoribonucleotides in aqueous solution are indispensable to study the recognition mechanisms of reader proteins and their cognate RNA binders. This is because an oligoribonucleotide ligand has much more degrees of freedom compared to a normal-size small-molecule ligand in aqueous solution.

The conformational analysis suggests a conformational selection model. In this model, we assume that the binding pocket of YTHDC1 (i.e., the aromatic cage) is relatively rigid with only one most populated conformation. This prevalent conformation is competent for recognizing the m⁶A-containing oligoribonucleotides in its apo form. During the conformational selection by the protein, those oligoribonucleotides motifs that mainly populate the bound state in aqueous solution will pay less penalty for the conformational change than those predominating bound-resistant conformations. This model explains why GG(m⁶A)CU binds stronger than G(m⁶A)CU, thus elucidating the role of 5' terminus G₋₂ in the binding. Our analysis also shows that m⁶A is mostly solvent exposed in the bound-like conformations (Figure S8), which is prerequisite for the recognition by the aromatic cage. In fact, the adenine also tends to be more solvent exposed in the methylated GGACU than in the unmethylated counterpart (Figure S8B). This suggests that the methyl group not only contributes the interactions of oligoribonucleotides with reader proteins but also facilitates the nucleotides adopting a bound-like conformation. The capacity of the methyl group for the conformational regulation is not only limited to explain the recognition mechanism for reader proteins. The exposed conformation of methylated adenine would be also required to accommodate it in the active site of m⁶A-specific demethylase FTO (X. Zhang et al., 2019).

Differential and shared recognition mechanisms between YTHDC1 and other reader proteins

The five human m⁶A reader proteins can be classified into two classes according to their binding pocket features, with YTHDC1 and YTHDC2 in one class and YTHDF1, YTHDF2 and YTHDF3 in the other (Figure S10). Here, we select YTHDC1 and YTHDF1 as representatives in the two classes. The first difference is in the side chain that interacts with the N1 position of m⁶A (Figure S10). Asn367 of YTHDC1 interacts with the N1 atom by donating a hydrogen atom, whilst the corresponding side chain in YTHDF1 is Asp401. The interaction between the N1 atom and the asparagine side chain is more favorable than aspartic acid (C. Xu et al., 2015). Second, the two reader proteins have different gate keeper residues at the border of their binding pockets, i.e., Asn363 and Tyr397 in YTHDC1 and YTHDF1, respectively. For the latter, the Tyr397Ala mutation abolishes its binding to GG(m⁶A)CU (C. Xu et al., 2015). The X-ray structure of holo YTHDF1 shows that Tyr397 forms π - π stacking with the 5' terminal nucleotides (C. Xu et al., 2015). In contrast to YTHDC1, such specific interaction probably restrains the 5' terminus dynamics on YTHDF1.

The two classes of reader proteins share two structural features, i.e., the aromatic cage for recognizing m⁶A and a positively charged shallow pocket for interactions with the 3' terminal nucleotides. The two conserved features suggest a common mechanism for the binding of reader proteins to a m⁶A-containing RNA. First, the aromatic cage permits the selective recognition of m⁶A. The indole side chain of one of the tryptophans provides a π - π stacking with the modified adenine while the other indole is at optimal van der Waals distance with the methyl group. Our simulations show that GG(m⁶A)CU tends to expose its m⁶A to the solvent in its free state comparing to its unmethylated counterpart (Figure S8B). The m⁶A exposure provides a specific geometry for fitting the aromatic cage in all reader proteins, and thereby distinguishing unmethylated RNAs. Second, the surface pocket adjacent to the aromatic cage conservatively contains two charged residues in all five reader proteins, i.e., Arg475 and Lys361 for YTHDC1 (Figure S10). Their importance for the binding of m⁶A-containing RNAs has been verified in several reader proteins (S. K. Luo & Tong, 2014; C. Xu et al., 2015; C. Xu et al., 2014). The 3' terminal nucleotides contribute to the protein-RNA binding by specific association to the positively charged pocket on the YTHDC1 surface. The structural features suggest a two-step binding mode, in which the electrostatically-driven association of the 3' terminal nucleotides facilitates the further recognition and formation of optimal contacts between m⁶A and the aromatic cage (Zhu et al., 2014). Such mechanism might be common to the five human m⁶A reader proteins as they share a positive electrostatic potential on the surface that recognizes the 3' terminus nucleotides (Figure S11). The inverse sequence of events

is less probable considering that electrostatic attraction has a longer range than van der Waals interactions.

Accession number

PDB IDs: 6RT4, 6RT5, 6RT6 and 6RT7.

Acknowledgements

The use of beamlines and user support at Swiss Light Source are gratefully acknowledged. We thank the Swiss National Supercomputing Centre (CSCS) in Lugano for providing the computational resources. We also thank Dr. Xiang Wang for valuable discussions.

Funding

This work was supported by the Swiss National Science Foundation (to A.C.), the International Postdoc Grant funded by the Swedish Research Council (to Y.L), and the grant from the Innosuisse (Swiss Innovation Agency) as well as Entrepreneur Fellowship in Biotechnology funded by University of Zurich (to P.Š).

References

- Aduri, R., Psciuk, B. T., Saro, P., Taniga, H., Schlegel, H. B., & SantaLucia, J. (2007). AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *Journal of Chemical Theory and Computation*, 3(4), 1464-1475. doi: 10.1021/ct600329w
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., . . . Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D-Structural Biology*, 68, 352-367. doi: 10.1107/S0907444912001308
- Alarcon, C. R., Lee, H., Goodarzi, H., Halberg, N., & Tavazoie, S. F. (2015). N6-methyladenosine marks primary microRNAs for processing. *Nature*, 519(7544), 482-485. doi: 10.1038/nature14281
- Banas, P., Hollas, D., Zgarbova, M., Jurecka, P., Orozco, M., Cheatham, T. E., . . . Otyepka, M. (2010). Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *Journal of Chemical Theory and Computation*, 6(12), 3836-3849. doi: 10.1021/ct100481h
- Banas, P., Jurecka, P., Walter, N. G., Sponer, J., & Otyepka, M. (2009). Theoretical studies of RNA catalysis: Hybrid QM/MM methods and their comparison with MD and QM. *Methods*, 49(2), 202-216. doi: 10.1016/j.ymeth.2009.04.007
- Batista, P. J., Molinie, B., Wang, J. K., Qu, K., Zhang, J. J., Li, L. J., . . . Chang, H. Y. (2014). m(6)A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell*, 15(6), 707-719. doi: 10.1016/j.stem.2014.09.019
- Ben-Haim, M. S., Moshitch-Moshkovitz, S., & Rechavi, G. (2015). FTO: linking m(6)A demethylation to adipogenesis. *Cell Res*, 25(1), 3-4. doi: 10.1038/cr.2014.162
- Boriack-Sjodin, P. A., Ribich, S., & Copeland, R. A. (2018). RNA-modifying proteins as anticancer drug targets. *Nature Reviews Drug Discovery*, 17(6), 435-453. doi: 10.1038/nrd.2018.71
- Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., . . . Karplus, M. (2009). CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*, 30(10), 1545-1614. doi: 10.1002/jcc.21287
- Changeux, J. P., & Edelstein, S. (2011). Conformational selection or induced fit? 50 years of debate resolved. *F1000 biology reports* 3, 19. doi: 10.3410/B3-19
- Dai, D. J., Wang, H. Y., Zhu, L. Y., Jin, H. C., & Wang, X. (2018). N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death & Disease*, 9. doi: 10.1038/S41419-017-0129-X
- Delaunay, S., & Frye, M. (2019). RNA modifications regulating cell fate in cancer. *Nature Cell Biology*, 21(5), 552-559. doi: 10.1038/s41556-019-0319-0
- Denning, E. J., Priyakumar, U. D., Nilsson, L., & Mackerell, A. D. (2011). Impact of 2'-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *Journal of Computational Chemistry*, 32(9), 1929-1943. doi: 10.1002/jcc.21777
- Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N., & Rechavi, G. (2013). Transcriptome-wide mapping of N-6-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nature Protocols*, 8(1), 176-189. doi: 10.1038/nprot.2012.148
- Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. F., & Shaw, D. E. (2012). Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics*, Vol 41, 41, 429-452. doi: 10.1146/annurev-biophys-042910-155245
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography*, 66, 486-501. doi: 10.1107/S0907444910007493
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics*, 103(19), 8577-8593. doi: 10.1063/1.470117

- Feller, S. E., Zhang, Y. H., Pastor, R. W., & Brooks, B. R. (1995). Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *Journal of Chemical Physics*, 103(11), 4613-4621. doi: 10.1063/1.470648
- Fustin, J. M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., . . . Okamura, H. (2013). RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell*, 155(4), 793-806. doi: 10.1016/j.cell.2013.10.026
- Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., . . . Hanna, J. H. (2015). m(6)A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science*, 347(6225), 1002-1006. doi: 10.1126/science.1261417
- Hsu, P. J., Shi, H. L., & He, C. (2017). Epitranscriptomic influences on development and disease. *Genome Biology*, 18. doi: 10.1186/S13059-017-1336-6
- Huang, J., & MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, 34(25), 2135-2145. doi: 10.1002/jcc.23354
- Jia, G. F., Fu, Y., Zhao, X., Dai, Q., Zheng, G. Q., Yang, Y., . . . He, C. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*, 7(12), 885-887. doi: 10.1038/Nchembio.687
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics*, 79(2), 926-935. doi: Doi 10.1063/1.445869
- Kabsch, W. (2010). Xds. *Acta Crystallographica Section D-Biological Crystallography*, 66, 125-132. doi: 10.1107/S0907444909047337
- Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules (vol 9, pg 646, 2002). *Nature Structural Biology*, 9(10), 788-788. doi: 10.1038/nsb1002-788a
- Krepl, M., Clery, A., Blatter, M., Allain, F. H. T., & Sponer, J. (2016). Synergy between NMR measurements and MD simulations of protein/RNA complexes: application to the RRM, the most common RNA recognition motifs. *Nucleic Acids Research*, 44(13), 6452-6470. doi: 10.1093/nar/gkw438
- Krepl, M., Havrila, M., Stadlbauer, P., Banas, P., Otyepka, M., Pasulka, J., . . . Sponer, J. (2015). Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein-RNA Complexes? *Journal of Chemical Theory and Computation*, 11(3), 1220-1243. doi: 10.1021/ct5008108
- Kuhrova, P., Best, R. B., Bottaro, S., Bussi, G., Sponer, J., Otyepka, M., & Banas, P. (2016). Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *Journal of Chemical Theory and Computation*, 12(9), 4534-4548. doi: 10.1021/acs.jctc.6b00300
- Kuhrova, P., Mlynsky, V., Zgarbova, M., Krepl, M., Bussi, G., Best, R. B., . . . Banas, P. (2019). Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J Chem Theory Comput*, 15(5), 3288-3305. doi: 10.1021/acs.jctc.8b00955
- Lemkul, J. A., & MacKerell, A. D. (2018). Polarizable force field for RNA based on the classical drude oscillator. *Journal of Computational Chemistry*, 39(32), 2624-2646. doi: 10.1002/jcc.25709
- Li, A., Chen, Y. S., Ping, X. L., Yang, X., Xiao, W., Yang, Y., . . . Yang, Y. G. (2017). Cytoplasmic m(6)A reader YTHDF3 promotes mRNA translation. *Cell Res*, 27(3), 444-447. doi: 10.1038/cr.2017.10
- Li, F. D., Zhao, D. B., Wu, J. H., & Shi, Y. Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. *Cell Res*, 24(12), 1490-1492. doi: 10.1038/cr.2014.153
- Lin, S. B., Choe, J., Du, P., Triboulet, R., & Gregory, R. I. (2016). The m(6)A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Molecular Cell*, 62(3), 335-345. doi: 10.1016/j.molcel.2016.03.021
- Liu, J. Z., Yue, Y. N., Han, D. L., Wang, X., Fu, Y., Zhang, L., . . . He, C. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature Chemical Biology*, 10(2), 93-95. doi: 10.1038/nchembio.1432

- Liu, N., Dai, Q., Zheng, G. Q., He, C., Parisien, M., & Pan, T. (2015). N-6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, 518(7540), 560-564. doi: 10.1038/nature14234
- Liu, N., & Pan, T. (2016). N-6-methyladenosine-encoded epitranscriptomics. *Nature Structural & Molecular Biology*, 23(2), 98-102. doi: 10.1038/nsmb.3162
- Luo, S. K., & Tong, L. (2014). Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proceedings of the National Academy of Sciences of the United States of America*, 111(38), 13834-13839. doi: 10.1073/pnas.1412742111
- Luo, Y., & Roux, B. (2010). Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *Journal of Physical Chemistry Letters*, 1(1), 183-189. doi: 10.1021/jz900079w
- Mackerell, A. D., & Nilsson, L. (2008). Molecular dynamics simulations of nucleic acid-protein complexes. *Current Opinion in Structural Biology*, 18(2), 194-199. doi: 10.1016/j.sbi.2007.12.012
- Martyna, G. J., Tobias, D. J., & Klein, M. L. (1994). Constant-Pressure Molecular-Dynamics Algorithms. *Journal of Chemical Physics*, 101(5), 4177-4189. doi: Doi 10.1063/1.467468
- Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658-674. doi: 10.1107/S0021889807021206
- Meyer, K. D., & Jaffrey, S. R. (2014). The dynamic epitranscriptome: N-6-methyladenosine and gene expression control. *Nature Reviews Molecular Cell Biology*, 15(5), 313-326. doi: 10.1038/nrm3785
- Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., & Jaffrey, S. R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell*, 149(7), 1635-1646. doi: 10.1016/j.cell.2012.05.003
- Musiani, F., Rossetti, G., Capece, L., Gerger, T. M., Micheletti, C., Varani, G., & Carloni, P. (2014). Molecular Dynamics Simulations Identify Time Scale of Conformational Changes Responsible for Conformational Selection in Molecular Recognition of HIV-1 Transactivation Responsive RNA. *Journal of the American Chemical Society*, 136(44), 15631-15637. doi: 10.1021/ja507812v
- Nam, K., Gao, J. L., & York, D. M. (2008). Electrostatic interactions in the hairpin ribozyme account for the majority of the rate acceleration without chemical participation by nucleobases. *RNA*, 14(8), 1501-1507. doi: 10.1261/rna.863108
- Nam, K. H., Gaot, J. L., & York, D. M. (2008). Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *Journal of the American Chemical Society*, 130(14), 4680-4691. doi: 10.1021/ja0759141
- Peng, L., Yuan, X. Q., Jiang, B. Y., Tang, Z. L., & Li, G. C. (2016). LncRNAs: key players and novel insights into cervical cancer. *Tumor Biology*, 37(3), 2779-2788. doi: 10.1007/s13277-015-4663-9
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781-1802. doi: 10.1002/jcc.20289
- Qi, S. T., Ma, J. Y., Wang, Z. B., Guo, L., Hou, Y., & Sun, Q. Y. (2016). N-6-Methyladenosine Sequencing Highlights the Involvement of mRNA Methylation in Oocyte Meiotic Maturation and Embryo Development by Regulating Translation in *Xenopus laevis*. *Journal of Biological Chemistry*, 291(44), 23020-23026. doi: 10.1074/jbc.M116.748889
- Robertson, M. J., Qian, Y., Robinson, M. C., Tirado-Rives, J., & Jorgensen, W. L. (2019). Development and Testing of the OPLS-AA/M Force Field for RNA. *Journal of Chemical Theory and Computation*, 15(4), 2734-2742. doi: 10.1021/acs.jctc.9b00054
- Salmon, L., Bascom, G., Andricioaei, I., & Al-Hashimi, H. M. (2013). A General Method for Constructing Atomic-Resolution RNA Ensembles using NMR Residual Dipolar Couplings: The

- Basis for Interhelical Motions Revealed. *Journal of the American Chemical Society*, 135(14), 5457-5466. doi: 10.1021/ja400920w
- Satterlee, J. S., Basanta-Sanchez, M., Blanco, S., Li, J. B., Meyer, K., Pollock, J., . . . Rybak-Wolf, A. (2014). Novel RNA Modifications in the Nervous System: Form and Function. *Journal of Neuroscience*, 34(46), 15170-15177. doi: 10.1523/Jneurosci.3236-14.2014
- Shi, H., Wang, X., Lu, Z., Zhao, B. S., Ma, H., Hsu, P. J., . . . He, C. (2017). YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA. *Cell Res*, 27(3), 315-328. doi: 10.1038/cr.2017.15
- Sledz, P., & Caflish, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology*, 48, 93-102. doi: 10.1016/j.sbi.2017.10.010
- Sledz, P., & Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. *Elife*, 5. doi: 10.7554/eLife.18434
- Sponer, J., Krepl, M., Banas, P., Kuhrova, P., Zgarbova, M., Jurecka, P., . . . Otyepka, M. (2017). How to understand atomistic molecular dynamics simulations of RNA and protein-RNA complexes? *Wiley Interdisciplinary Reviews: RNA*, 8(3). doi: 10.1002/Wrna.1405
- Steinbach, P. J., & Brooks, B. R. (1994). New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *Journal of Computational Chemistry*, 15(7), 667-683. doi: DOI 10.1002/jcc.540150702
- Tan, D., Piana, S., Dirks, R. M., & Shaw, D. E. (2018). RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proceedings of the National Academy of Sciences of the United States of America*, 115(7), E1346-E1355. doi: 10.1073/pnas.1713027115
- Templeton, C., & Eber, R. (2018). Why Does RNA Collapse? The Importance of Water in a Simulation Study of Helix-Junction-Helix Systems. *Journal of the American Chemical Society*, 140(49), 16948-16951. doi: 10.1021/jacs.8b11111
- Vangaveti, S., Ranganathan, S. V., & Chen, A. A. (2017). Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *Wiley Interdisciplinary Reviews: RNA*, 8(2). doi: 10.1002/wrna.1396
- Vermeulen, M. (2017). N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Molecular & Cellular Proteomics*, 16(8), S42-S42.
- Wang, P., Duxtader, K. A., & Nam, Y. (2016). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Molecular Cell*, 63(2), 306-317. doi: 10.1016/j.molcel.2016.05.041
- Wang, X., Feng, J., Xue, Y., Guan, Z. Y., Zhang, D. L., Liu, Z., . . . Yin, P. (2017). Structural basis of N6-adenosine methylation by the METTL3-METTL14 complex (vol 534, pg 575, 2016). *Nature*, 542(7640), 260-260. doi: 10.1038/nature21073
- Wang, X., Lu, Z. K., Gomez, A., Hon, G. C., Yue, Y. N., Han, D. L., . . . He, C. (2014). N-6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 505(7481), 117-120. doi: 10.1038/nature12730
- Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z. K., Han, D. L., Ma, H. H., . . . He, C. (2015). N-6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*, 161(6), 1388-1399. doi: 10.1016/j.cell.2015.05.014
- Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z. L., & Zhao, J. C. (2014). N-6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature Cell Biology*, 16(2), 191-198. doi: 10.1038/ncb2902
- Wiedmer, L., Eberle, S. A., Bedi, R. K., Sledz, P., & Caflish, A. (2019). A Reader-Based Assay for m(6)A Writers and Erasers. *Anal. Chem.*, 91(4), 3078-3084. doi: 10.1021/acs.analchem.8b05500
- Wojtas, M. N., Pandey, R. R., Mendel, M., Homolka, D., Sachidanandam, R., & Pillai, R. S. (2017). Regulation of m(6) A Transcripts by the 3' -> 5' RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. *Molecular Cell*, 68(2), e26116. doi: 10.1016/j.molcel.2017.09.021

- Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., . . . Yang, Y. G. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell*, 61(4), 507-519. doi: 10.1016/j.molcel.2016.01.012
- Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., & Min, J. R. (2015). Structural Basis for the Discriminative Recognition of N-6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *Journal of Biological Chemistry*, 290(41), 24902-24913. doi: 10.1074/jbc.M115.680389
- Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., . . . Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nature Chemical Biology*, 10(11), 927-929. doi: 10.1038/nchembio.1654
- Xu, Y., MacKerell, A. D., & Nilsson, L. (2016). Structural effects of modified ribonucleotides and magnesium in transfer RNAs. *Bioorganic & Medicinal Chemistry*, 24(20), 4826-4834. doi: 10.1016/j.bmc.2016.06.037
- Xu, Y., Vanommeslaeghe, K., Aleksandrov, A., MacKerell, A. D., & Nilsson, L. (2016). Additive CHARMM Force Field for Naturally Occurring Modified Ribonucleotides. *Journal of Computational Chemistry*, 37(10), 896-912. doi: 10.1002/jcc.24307
- Yoo, J., & Aksimentiev, A. (2016). Improved Parameterization of Amine-Carboxylate and Amine-Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *Journal of Chemical Theory and Computation*, 12(1), 430-443. doi: 10.1021/acs.jctc.5b00967
- Yoo, J., & Aksimentiev, A. (2018). New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Physical Chemistry Chemical Physics*, 20(13), 8432-8449. doi: 10.1039/c7cp08185e
- Yoo, J. J., & Aksimentiev, A. (2012). Improved Parametrization of Li⁺, Na⁺, K⁺, and Mg²⁺ Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *Journal of Physical Chemistry Letters*, 3(1), 45-50. doi: 10.1021/jz201501a
- Yuan, Y. N., Zhang, Z. Z., Mills, M. J. L., Hu, R. J., & Zhang, R. S. (2018). Assessing Force Field Potential Energy Function Accuracy via a Multipolar Description of Atomic Electrostatic Interactions in RNA. *Journal of Chemical Information and Modeling*, 58(11), 2239-2254. doi: 10.1021/acs.jcim.8b00328
- Zhang, C. S., Lu, C., Jing, Z. F., Wu, C. J., Piquemal, J. P., Ponder, J. W., & Ren, P. Y. (2018). AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *Journal of Chemical Theory and Computation*, 14(4), 2084-2108. doi: 10.1021/acs.jctc.7b01169
- Zhang, X., & Jia, G. F. (2016). RNA epigenetic modification: N6-methyladenosine. *Yi Chuan*, 38(4), 275-288. doi: 10.16288/j.ycz.16-049
- Zhang, X., Wei, L. H., Wang, Y. X., Xiao, Y., Liu, J., Zhang, W., . . . Jia, G. F. (2019). Structural insights into FTO's catalytic mechanism for the demethylation of multiple RNA substrates. *Proceedings of the National Academy of Sciences of the United States of America*, 116(8), 2919-2924. doi: 10.1073/pnas.1820574116
- Zheng, G. Q., Dahl, J. A., Niu, Y. M., Fedorcsak, P., Huang, C. M., Li, C. J., . . . He, C. (2013). ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Molecular Cell*, 49(1), 18-29. doi: 10.1016/j.molcel.2012.10.015
- Zhou, J., Wan, J., Gao, X. W., Zhang, X. Q., Jaffrey, S. R., & Qian, S. B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature*, 526(7574), 591-U332. doi: 10.1038/nature15377
- Zhu, T. T., Roundtree, I. A., Wang, P., Wang, X., Wang, L., Sun, C., . . . Xu, Y. H. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res*, 24(12), 1493-1496. doi: 10.1038/cr.2014.152