Recognition of N6-methyladenosine by the YTHDC1 YTH domain studied by molecular dynamics and NMR spectroscopy: The role of hydration

Miroslav Krepl,[1]* Fred Franz Damberger,[2] Christine von Schroetter,[2] Dominik Theler,[2] Pavlína Pokorná,[1,3] Frédéric H.-T. Allain,[2] and Jiří Šponer[1]

[1]Institute of Biophysics of the Czech Academy of Sciences, Kralovopolska 135, 612 65 Brno, Czech Republic

[2]Department of Biology, ETH Zürich, Institute of Biochemistry, 8093 Zürich, Switzerland

[3]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

* Corresponding author, email: krepl@seznam.cz

Abstract

Background: The YTH domain of YTHDC1 belongs to a class of protein "readers", recognizing the N6-methyladenosine (m^6A) chemical modification in mRNA. Static ensemble-averaged structures revealed details of N6-methyl recognition via a conserved aromatic cage.

Methods: We performed molecular dynamics (MD) simulations along with nuclear magnetic resonance (NMR) and isothermal titration calorimetry (ITC) measurements to examine how dynamics and solvent interactions contribute to the m^6A recognition and negative selectivity towards unmethylated substrate.

Results: An intricate network of water-mediated interactions surrounds bound m^6A. The unmethylated adenosine allows disruptive intrusions of bulk water deep into the binding pocket, increasing selectivity for m^6A. We furthermore show that the YTHDC1's preference for the 5′-Gm^6A-3′ motif is partially facilitated by a network of water-mediated interactions between the 2-amino group of the preceding guanosine and residues deep in the m^6A binding pocket. The 5′-Im^6A-3′ (where I is inosine) motif can be recognized as well at the cost of disrupting the water network and a small decrease in affinity. The YTHDC1 D479A mutant, which interrupts the water network, also destabilizes m^6A binding. Lastly, we formulate and test an easy-to-implement approach for increasing the agreement between simulations and NMR experiments by using the HBfix potential function for stabilization of key NOE distances. We call the new approach NOEfix.

Conclusions: The structured water molecules surrounding the bound RNA and the methylated substrate's ability to exclude bulk water molecules are important elements of the YTH domain's preference for m^6A. Network of water molecules also fine tunes the specificity towards 5′-Gm^6A-3′ motifs.

General Significance: Our interdisciplinary study of YTHDC1 protein/RNA complex reveals an unusual mechanism by which solvent dynamics can contribute towards recognition of methylation by proteins.

Keywords: YTH domain, YTHDC1, water-bridges, hydration, molecular dynamics, N6-methyladenosine, NMR spectroscopy.

1. Introduction

N6-methyladenosine (m^6A) is the most common posttranscriptional modification of nucleobases in mRNA [1, 2]. It can influence basic cellular processes such as cell differentiation and proliferation by changing the stability of specific RNA transcripts or by promoting their translation [3, 4]. It is most often localized near stop-codons or in 5′- or 3′-untranslated regions of mRNA [5]. To handle the full cycle of the m^6A metabolism, the cell requires three types of proteins that can introduce, detect, and remove this modification [6-8]. These proteins are commonly known as writers, readers and erasers, respectively. Among the readers, the YT521-B homology (YTH) family of proteins is the most prominent. The YTH domain is evolutionary conserved in eukaryotes, possessing a characteristic $\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3\beta_4\beta_5\beta_6\alpha_4$ secondary structure [3, 9]. It facilitates recognition of the methyl group in N6-methyladenosine by a set of aromatic amino acids, including two tryptophans which are absolutely conserved in all YTH domains [10]. These residues form an aromatic cage which binds N6-methyladenosine with 50-fold higher affinity than adenosine (Figure 1). It was suggested that this preference is based on a set of van der Waals (vdW) interactions between the residues lining the m^6A binding pocket (including the conserved tryptophans) and the N6-methyl group [10]. Structurally, the aromatic cage is similar to the binding sites of methylated lysine observed in the evolutionary unrelated chromo and tudor protein domains [11], making it a good example of convergent evolution in recognition of methylated modifications of biomolecules, be they proteins or nucleic acids. Here, we explore the YTH domain of YTHDC1 protein which facilitates the regulation of pre-mRNA splicing by recruiting splicing factors [12]. The YTHDC1 protein interacts with RNA solely by its single YTH domain [9]. Unusually, the YTHDC1 preferentially recognizes 5’-Gm^6A-3’ motifs in RNA [13, 14]. The other YTH-domain-containing proteins, such as the YTHDF family [15], do not display any preference for nucleotide preceding m^6A.

We use molecular dynamics (MD) simulations to study the YTH domain of the YTHDC1 protein, both in isolation and bound to RNA oligonucleotide containing the 5’-Gm^6A-3’ motif (Figure 1). MD describes the Boltzmann distribution (populations) of molecular conformations using carefully calibrated empirical models – molecular mechanical force fields. In the past, the MD simulations have been successfully applied many times in studies of protein/RNA complexes [16], including systems containing the YTH domain [14, 17, 18]. By providing essentially infinite temporal and spatial resolution, the MD can extend upon the information from ensemble-averaged experimental data. It is particularly well-suited for studies of hydration and its structural role in biomolecular complexes [19, 20]. The main limitations of MD are the quality of the force field and the affordable timescale (sampling) of the simulation [16], necessitating careful interpretation of the simulation data. Thus we apply NMR spectroscopy experiments and isothermal titration calorimetry (ITC) measurements to verify and corroborate the most important results predicted by our simulations [21].

By elucidating the role of dynamics and solvent interactions, our data provides a more complete picture of the YTH domain preference for m^6A over A, going beyond the static structural details revealed by experimental data [13, 22, 23]. We show that in addition to the absence of stabilizing vdW interactions with the protein, the unmethylated adenosine is also unable to fully displace bulk water molecules from the m^6A binding pocket which subsequently disrupts its binding and can serve to facilitate fast turnover of incorrect

substrates. The unique preference of YTHDC1 for the 5'-Gm^{6}A-3' motif is primarily facilitated by direct protein/RNA H-bonds with the guanosine (henceforth referred to as G_2), as shown by the experimental structures. [13, 24] Here, our calculations reveal the G_2 is also involved in two guanine-specific water-bridge interactions via its 2-amino group. These water-bridge interactions connect the G_2 to a water network within the m^{6}A binding pocket. Mutating the G_2 into I_2 (inosine) abolishes this connection, impacting residues deep in the m^{6}A binding pocket and slightly reducing the protein/RNA binding affinity. We quantify the overall change in free energy of RNA binding by thermodynamic integration (TI) free-energy calculations and ITC experimental measurements and find good agreement. The water network can be even more disrupted by a D479A mutation in the protein.

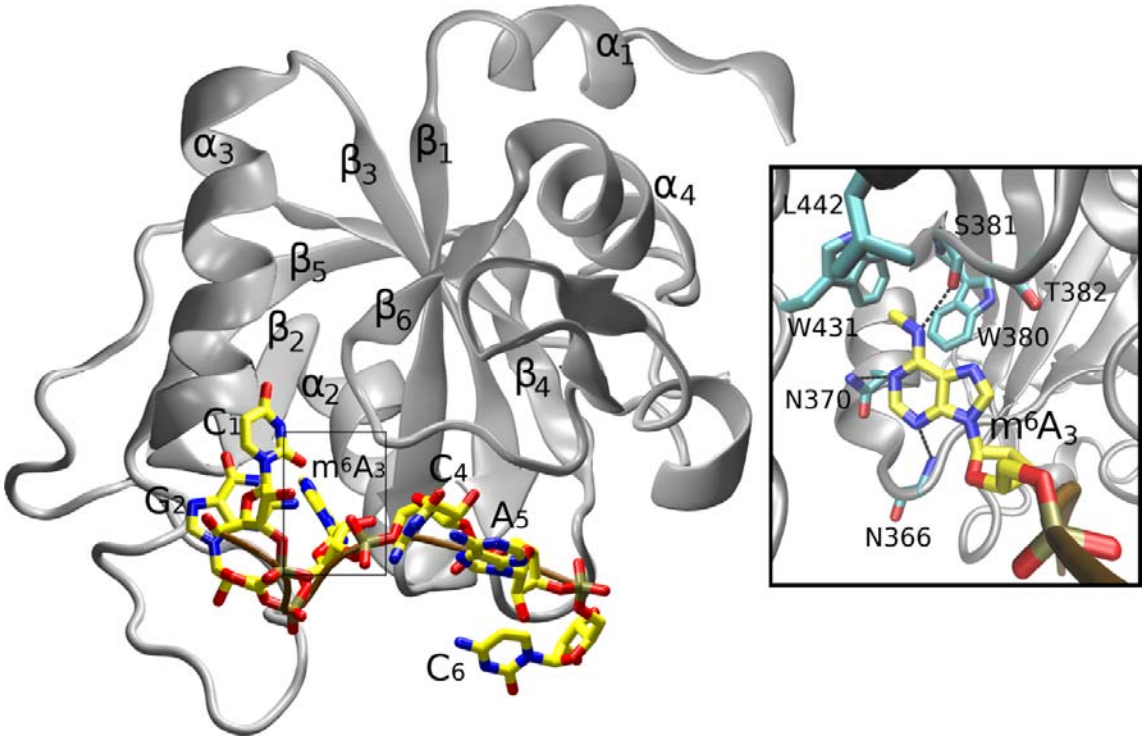

Figure 1. YTH domain of the YTHDC1 protein complexed with a 5′-$C_1G_2m^6A_3C_4A_5C_6$-3′ RNA oligonucleotide [13]. Detail of the m^{6}A binding pocket is shown in the inset. Secondary structure of the protein is labeled and shown as gray ribbon. The RNA backbone is in brown. The carbon atoms are colored yellow and cyan in RNA and protein, respectively, whereas nitrogen, oxygen and phosphorus are blue, red, and gold. Protein/RNA H-bonds are marked with dashed black lines. Unless specified otherwise, same color coding of atoms and marking for H-bonds were used in the subsequent figures.

2. Material and Methods

2.1. System building.

We used structures of the free YTHDC1 (PDB: 4r3h, 1.9 Å resolution X-ray structure) [22] and its complex with 5′-CGm^{6}ACAC-3′ RNA (PDB: 2mtv, first conformer of the NMR structure) [13] as starting structures for all our MD simulations. The initial files were prepared by the xLeap module of AMBER 18 [25]. All experimentally determined atoms were utilized. The missing residue side-chains in the structure of the free YTHDC1 were added by xLeap and the structure of the missing β4/β5 protein loop was taken from another structure of the YTHDC1 complexed with RNA (PDB: 4r3i, 1.8 Å resolution X-ray structure) [22]. Both the 4r3i and 2mtv structures contain 5'-Gm^{6}A-3' motif in the bound oligonucleotide and show a virtually identical protein/RNA interface. For our simulations of the protein/RNA complex, we have decided to utilize the 2mtv structure which was more convenient for this study (see Supplementary for details). We used the ff12SB [26] and bsc0χ_{OL3} (i.e., OL3) [27, 28] force fields to describe the protein and RNA, respectively. The electrostatic potentials of the N6,N1-dimethylated adenine base and of the methylamine were computed by Gaussian 09 [29] using HF/6-31G* level of theory. The partial charges of the m^{6}A base and of the methylamine were subsequently derived via the RESP method [30]. The remaining force-field parameters of m^{6}A were taken from the standard OL3 RNA force field and the parameters of the methylamine were adapted from parameters of the N6-methyl segment of the base. The partial charges and parameters of inosine were taken from Ref. [31] and OL3 parameters for guanosine [28] were used to describe its χ dihedral. In all simulations, the biomolecules were surrounded in an octahedral box of SPC/E [32] water molecules with minimal distance of 12 Å between the solute and the box border. Some simulations were also done with OPC waters and modified vdW parameters of phosphates supplemented by modified RNA backbone dihedrals (abbreviated as χ_{OL3}_CP(OPC) force-field combination) [33-35]. KCl ions [36] were added to neutralize all systems and to achieve an excess salt concentration of 0.15 M.

2.2. Simulation protocol.

The initial minimization and equilibration of the systems was performed in the pmemd.MPI implementation of AMBER according to the protocol described in Ref. [37]. In simulations of the protein/RNA complex structure, this was followed by 120 ns of simulations where we gradually released the experimental NMR-restraints. This protocol was designed to prevent abrupt departures from experimental geometry which is known to sometimes occur in early stages of simulations of NMR structures, as described in detail elsewhere [38]. Afterwards, the production simulations were run using the GPU-based pmemd.cuda [39] for 1 μs. In the preliminary set of simulations, we observed reversible perturbations of the native protein/RNA interactions for the C_4, A_5, and C_6 which, in some cases, led to loss of the protein/RNA interface for these peripheral nucleotides, but not the G_2 and m^{6}A$_3$ which were of our primary interest. Nevertheless, to avoid the potential sampling uncertainties when interpreting the data for the G_2 and m^{6}A$_3$ in such trajectories, we used mild 1 kcal/mol HBfix potential [40] to increase stability of the native A_5(OP1)/K364(NZ) and C_6(OP1)/K475(NZ) interactions in all subsequent simulations. This entirely prevented any problems on our simulation timescale. See Supplementary for further details. In all simulations, we used the SHAKE algorithm [41] and hydrogen mass repartitioning [42], allowing a 4 fs integration step. Particle mesh Ewald summation [43] was used to describe electrostatic interactions and

periodic boundary conditions were applied to handle system border bias. The cut-off distance for Lennard-Jones interactions was 9 Å. A Langevin thermostat and a Monte Carlo barostat [25] were used to regulate temperature and pressure, respectively.

2.3. Thermodynamic integration.

We used a mixed single/double topology approach as implemented in AMBER 18 (including update 16) to set up the Thermodynamic integration (TI) calculations. Specifically, a segment of the mutated residue undergoing the alchemical transformation was represented by a dual topology model while the rest of the molecule was described by single topology. We used the soft-core vdW potentials [44] to describe the appearing and disappearing atoms in the dual topology region. The simulations were carried out for nine lambda windows in parallel runs. The length of the individual lambda windows was 20 and 100 ns for reference RNA single-strand and protein/RNA complex simulations, respectively. The length of the complex simulations was greater to properly sample the conformational and water-network changes at the protein/RNA interface induced by the alchemical transformation. See Supplementary for further details.

2.4. Protein and RNA Preparation.

The ^{15}N-labeled YTH domain of *R. Norvegicus* YTHDC1 (residues 347- 502) was prepared as previously described [13] except that for the size exclusion chromatography (SEC) a buffer consisting of 25 mM NaH_2PO_4, 25 mM NaCl and 10 mM β-Mercaptoethanol at pH 7 (pH adjusted with NaOH) was used and 880 units per ml SEC sample RNAseOUT(Invitrogen) as RNAse inhibitor were added to the sample prior to the size exclusion chromatography purification step. Note that in Ref. [13], it was erroneously reported that the SEC buffer was prepared using NaH_2PO_4 instead of Na_2HPO_4. RNAs were obtained from Horizon Discovery Ltd, deprotected according to the supplier's instructions, lyophilized and resuspended in the SEC buffer.

2.5. Isothermal titration calorimetry (ITC).

Protein and RNA were dialyzed in SEC buffer before ITC measurements. ITC measurements were conducted at 20°C on a VP-ITC calorimeter (MicroCal) with a protein concentration of 10 µM in the cell and with syringe concentrations of 117 µM UGm^{6}ACAC and 124 µM UIm^{6}ACAC, respectively. An initial injection of 2 µl was followed by 45 injections of 6 µl each with an interval of 300 s between injections and a stirring speed of 307 rpm. Data analysis was perfomed with the Origin 7.0 software.

2.6. NMR.

NMR measurements were conducted in the SEC buffer supplemented with 10% D_2O as lock substance. NMR spectra were obtained on an AVIII 600MHz spectrometer equipped with a cryoprobe at 293K. Spectra were processed using Topspin3 (Bruker) and analysed with CARA [45] (www.nmr.ch). Titrations with RNA were performed by adding concentrated aliquots in NMR buffer to 0.9 mM YTHDC1 samples which showed slow exchange behavior

with separate signals observed for the free and RNA-bound states for both $I_2m^6A_3$ and $G_2m^6A_3$ RNAs. Assignments for the amides of YTHDC1 in the $I_2m^6A_3$ complex were obtained from a 3D ^{15}N-resolved NOESY of the complex by adjusting the position of H,N correlations so that NOESY towers matched the patterns observed in the $G_2m^6A_3$ complex. This was possible thanks to the highly similar aliphatic ^{1}H chemical shifts and NOESY towers in both structures. Final H,N peak positions used for obtaining chemical shift perturbations (CSPs) were then adjusted to match ^{15}N,^{1}H-HSQCs of the 1:1 $G_2m^6A_3$ and $I_2m^6A_3$ complexes. CSPs were calculated with the formula $CSP=[(0.2\Delta\delta_N)^2+\Delta\delta_H{}^2]^{1/2}$. Where $\Delta\delta_N$ and $\Delta\delta_H$ are the chemical shift differences of the two complexes for backbone amide ^{15}N and ^{1}H resonances, respectively.

2.7. Analyses.

We have used cpptraj and VMD for ensemble analysis and visual inspection of the MD simulation trajectories, respectively [46, 47] Raster3D, Povray and Gnuplot were used to prepare figures and graphs [48]. H-bonds were analyzed by monitoring atom distances between heavy atom donors and acceptors and donor-hydrogen-acceptor angles. The force-field performance of the protein/RNA complex simulations was evaluated by computing the $1/r^6$ averaged distance violations of the NOEs in the simulation ensemble by sander and comparing them with experimentally measured upper bounds [38]. The NOE distance in MD simulations was considered satisfied when its average value did not exceed the experimental upper bound distance plus 0.3 Å.

3. Results

3.1. General summary

We report a series of MD simulations (Table 1) of the YTH domain of YTHDC1 bound to m-^{6}A or standard adenosine RNA oligonucleotides to examine how dynamics in the complex is affected by the N6-methyl modification. We also simulated the free protein to characterize structure and dynamics of the m^6A binding pocket when no nucleotide is bound. The simulations with m^6A_3 revealed a network of structured water molecules and we subsequently performed simulations with an I_2 nucleotide present instead of G_2 to show how loss of the water interactions coordinated by the guanine's amino group affects the protein/RNA interface. The effects of binding I_2 or G_2-containing RNAs to YTH are also examined by NMR and ITC measurements. Lastly, we examine the dynamics of the m^6A_3 RNA substrate complexed with YTH D479A mutant which severely disrupts the water interaction network.

In addition, we have performed a number of simulations where we for the first time utilized the HBfix function to treat the experimental NOE distances measured by NMR. HBfix was originally suggested as a simple short force-field term that introduces a gentle stabilizing (or destabilizing) potential between two atoms with non-zero forces limited to a distance range relevant for H-bond interaction [40]. It can, for example, mimic an induction term. While the original HBfix has been suggested as a structure-specific term, it has been later used for overall tuning of the AMBER RNA force field; for more details, see Refs. [16,

49]. The HBfix functions can be easily modified to formally accommodate the NOE upper-bound distances instead of H-bonds. We call this approach "NOEfix" and suggest that its judicial application can increase agreement between the NMR data and MD simulations without suppressing conformational transitions and sampling of substates. NOEfix, in contrast to restraints, does not introduce a continuously growing penalty with increasing distances. In other words, it allows smooth sampling of substates that temporarily or even permanently violate NOEs, depending on the magnitude of the potential energy tuning. Although not a panacea, NOEfix potential offer an alternative and easily applicable tool for curation of MD simulation trajectories based on the NMR data.

Table 1. List of simulations.[a]

simulated system	**number of simulations × length [μs]**
2mtv_m^6A_3	8 × 1 μs
2mtv_m^6A_3_NOEfix312[b]	6 × 1 μs
2mtv_m^6A_3_NOEfix18[b]	6 × 1 μs
2mtv_ m^6A_3_CP[c]	5 × 1 μs
2mtv_A_3	4 × 1 μs
2mtv_I_2	4 × 1 μs
2mtv_D479A_m^6A_3	4 × 1 μs
4r3h	4 × 1 μs
4r3h_methylamine	2 × 1 μs
2mtv_TI_ m^6A_3/A_3	54 × 100 ns, 54 × 20 ns
2mtv_TI_ G_2/ I_2	54 × 100 ns, 54 × 20 ns

[a]Experimental structures of the free protein (PDB: 4r3h) [22] or the protein/RNA complex (PDB: 2mtv) [13] served as starting structures in simulations. The m^6A_3 and A_3 indicate that the N6-methyladenosine and standard adenosine, respectively, were the third RNA nucleotide. I_2 indicates a G to I substitution of the second nucleotide. D479A indicates that the aspartate 479 of YTH was mutated into alanine. Methylamine indicates that twenty methylamines were added into the simulation water box before start. [b]NOEfix function was applied to all (312) or selected (18) experimental NOE distances showing ensemble-averaged distance violations in the 2mtv_m^6A_3 simulations. [c]Modified vdW parameters of the phosphates [34], modified RNA backbone torsion angles [35], and OPC water model [33] were used.

3.2. The very stable interaction of N6-methylated adenosine with YTHDC1 is mediated by H-bonding, stacking and water bridges.

The experimental structure used as the start [13] shows m^6A_3 base forming stacking and H-bonding interactions with the protein. Namely, there are m^6A_3(N6)/S381(O), m^6A_3(N1)/N370(ND2), and m^6A_3(N3)/N366(N) H-bonds. In addition, one face of the base and the N6-methyl group stack with W380 and W431 side-chains, respectively, and the other face of the base stacks with the side-chain of L442 (Figure 1). These interactions were entirely stable in all simulations. In addition, we observed a network of structured water molecules (water bridges) facilitating additional protein/RNA interactions. Namely, there was a highly occupied water bridge involving four solute atoms – m^6A_3(N7), W380(NE1), T382(OG1), and D479(OD) (Figure 2), having water exchange time with bulk on the scale of hundreds of nanoseconds (Table 2). This hydration site is in excellent agreement with the

corresponding X-ray structure (PDB: 4r3i) [22]. The simulations also revealed a water bridge between T382(OG1) and D479(OD) and another between m^6A_3(O2') and K364(O) (Figure 2). These water molecules are not visible in the X-ray structure and were thus uniquely predicted by the MD simulations with an exchange timescale of tens of nanoseconds. Overall, the simulations predict an intricate network of water molecules within and near the entrance to the m^6A_3 binding pocket (Figure 2). The NMR experimental structure used as the starting structure in our protein/RNA simulations does not provide any *a priori* information about the distribution of solvent molecules and all observed hydration sites were thus spontaneously established during the equilibration or early stages of the simulations. Notably, we observed the same water network near the m^6A_3 binding pocket also while using the OPC water model (see Methods) so the results do not depend on the utilized water model.

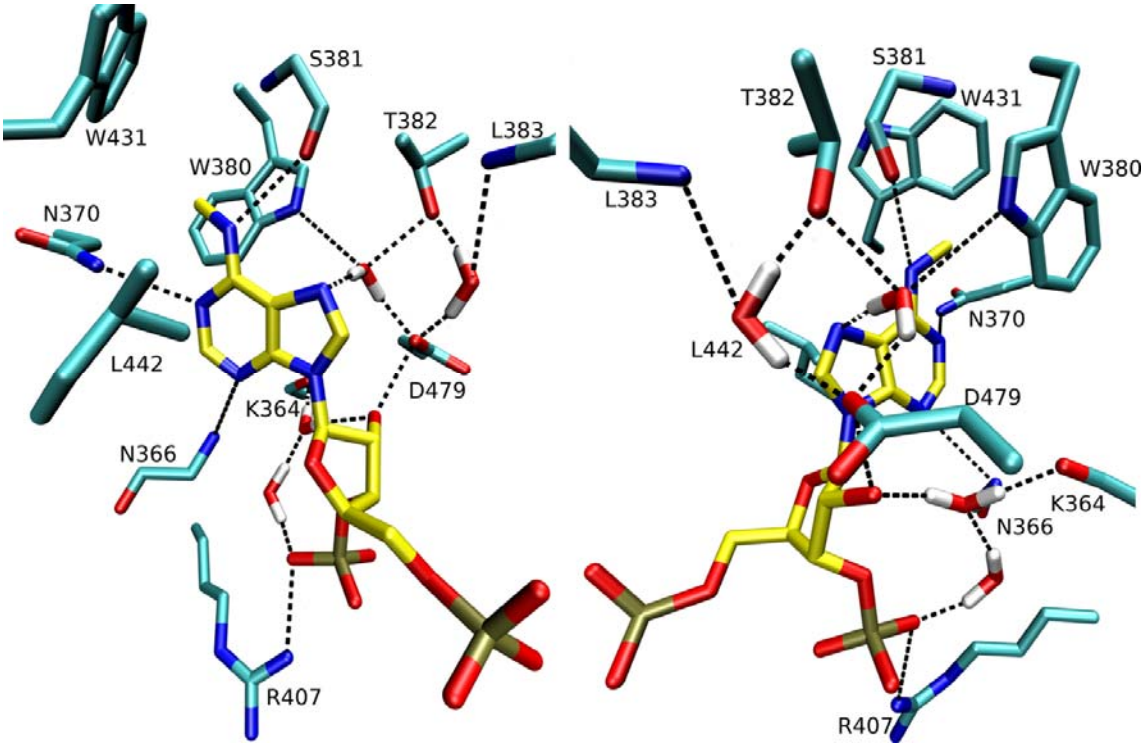

Figure 2. Recognition of the m^6A_3 nucleotide in MD simulations shown from two different angles. The participating amino-acids are labeled. Solvent hydrogens are white. For clarity, the solute hydrogen atoms are not displayed.

3.3. Unmodified adenosine permits invasions of water molecules into the m^6A_3 binding pocket.

Earlier experiments have reported ~50 fold stronger binding (free-energy difference of ~2.3 kcal/mol) of m^6A_3 to the YTH domain compared to A_3 [13]. A slightly higher value is indicated by our TI calculations (~3.6 ± 0.9 kcal/mol, see Supplementary). A recent study indicated 1.7 kcal/mol penalty although this was computed for binding of an isolated m^6A nucleoside rather than a full oligonucleotide [18]. Yet, highly similar NMR chemical shifts of the protein atoms were reported in the presence of both m^6A methylated and unmethylated RNAs, suggesting that they are bound identically [13]. To reconcile these observations in a structural context, we have performed simulations of the protein/RNA complex with

unmodified A_3 (by substituting the N6-methyl group with hydrogen in the starting structure). These simulations at first showed an identical development as for the m^6A_3 variant, with all H-bonds and stacking interactions preserved, except those formed by the N6-methyl group. The same water bridges near the binding pocket were formed, albeit with altered lifetimes (Table 2). However, all simulations subsequently showed periodic intrusions of a single water molecule from outside the pocket which could temporarily enter the region between the adenosine and the side-chain of W431 (Figure 3A). On average, a water molecule was present at this location in ~9% of the simulation frames. The same location is occupied by N6-methyl group in the m^6A_3 systems (Figure 2). We interpret these water intrusions to be a direct consequence of the methyl group's removal, resulting in a void which cannot be easily eliminated by reorientation of atoms forming the binding pocket. During its intrusions, the single water molecule formed one H-bond with the H61 atom of A_3 (Figure 3A) but since there were no other hydrogen bond donors or acceptors nearby, it was unstable in this position and was soon expelled again from the pocket. Eventually, another water molecule would come in, repeating the process. On average, in our simulation set we observed ~90 water molecule exchanges per 1 μs in this position.

In two simulations, we also observed that instead of being expelled, the single water molecule gradually attracted additional water molecules from the bulk solvent into the binding pocket (Figure 3B). In the two simulations where this happened, the distance between A_3 and the W431 side-chain started to increase as a result and the adenosine gradually drifted out of its binding pocket in order to make room for the incoming water molecules. In the first of the two simulations, this process was reversible on our simulation timescale. In the second, we observed a complete expulsion of A_3 from its binding pocket facilitated by these invading water molecules after ca 800 ns of simulation.

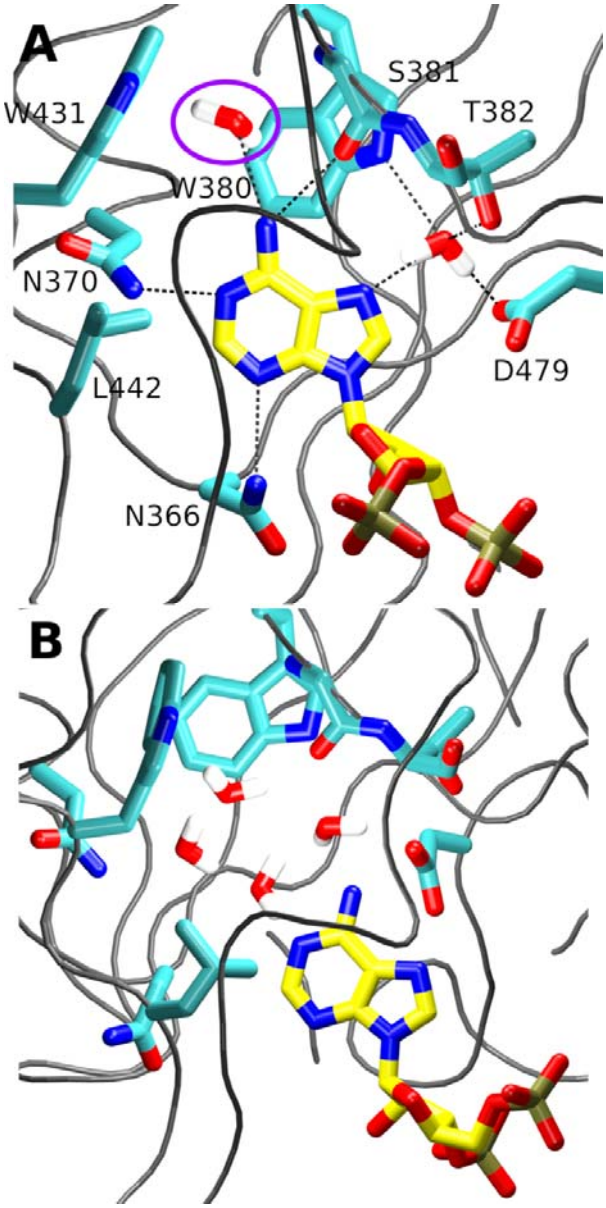

Figure 3. (**A**) A single water molecule (violet circle) periodically invaded the binding pocket of the A_3 nucleotide in MD simulations of the 2mtv_A_3 system, occupying the space where the N6-methyl group of the m^6A_3 nucleotide would normally be located. (**B**) In two simulations, the intrusion of the first water molecule attracted additional waters, in one case leading to expulsion of the A_3 nucleotide from its binding pocket.

3.4. The m^6A_3/A_3 binding pocket is filled with waters in simulations of the free protein.

Since unmethylated adenosine was unable to prevent water molecules from entering the binding pocket (see above and Figure 3), we hypothesized that it could also be less effective in displacing the water molecules present inside the binding pocket during the RNA binding. The need to displace these water molecules was supported by our simulations of the free protein (Table 1) [22] which showed that even though the side-chains forming the m^6A_3 binding pocket exhibited larger fluctuations than in the complex, the pocket was essentially

pre-formed in the free protein. It was also filled with water molecules at all times, with rapid sub-ns exchange with the bulk solvent (Figure 4). This indicates that during RNA binding, the incoming nucleotide must displace these water molecules from within the pocket.

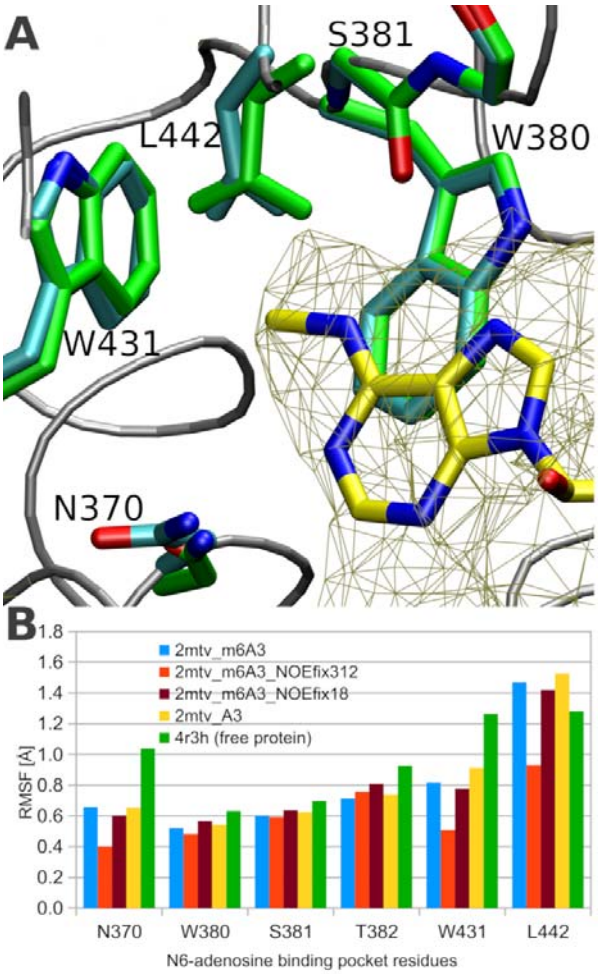

Figure 4. (**A**) Overlay of the average structures from 2mtv_m^6A_3 (protein/RNA complex; cyan carbons) and 4r3h (free protein; green carbons) simulations, respectively. The protein backbone and the bound m^6A_3 from the complex are shown in gray tube and with yellow carbons, respectively. In 4r3h simulations, the m^6A_3 binding pocket was composed and fully accessible for solvent molecules. The density grid of water molecules in the 4r3h simulations is indicated by the dark yellow mesh grid, overlapping with the binding site of m^6A_3 in the complex. (**B**) Graph of average per-residue root mean square fluctuation (RMSF) around the average simulation structure of selected systems, for residues lining the m^6A_3 binding pocket. With exception of L442, the range of fluctuations was greatest in simulations of the free protein.

3.5. Methylamine probe mimics binding of the m^6A to YTH.

In light of the above results, we wished to directly explore the binding process of the m^6A_3 nucleotide within the YTH binding pocket and the associated solvent dynamics. However, in

completely unbiased simulations, such process is unlikely to be observed on an affordable simulation timescale. For instance, in all our 2mtv_m^6A_3 simulations, we did not observe any signs of unbinding for the m^6A_3 nucleotide. A common approach in computational studies of substrate (un)binding is to induce or speed up the process by applying force potential bias along a single or multiple spatial coordinates, known as collective variables [50]. However, results of such calculations can depend on the chosen collective variables, as they represent prior information (reaction coordinate) of the process with unavoidable dimensionality reduction [16]. Here, the sampling of such process is further complicated by involvement of the bulk water molecules. We faced such sampling limitations in our free-energy TI calculations investigating unbinding of the methyl group in the framework of the thermodynamic cycle (see Supplementary). Thus, as an alternative approach to visualize accommodation of the methyl group in the binding pocket in the context of unbiased MD simulations, we have constructed a system in which we added twenty methylamine molecules into the water box surrounding the free YTH domain (Table 1). The methylamine is a structural mimic of the methylated amino group of the N6-methylated adenine. In the subsequent simulations, we have observed methylamine spontaneously binding in the m^6A_3 binding pocket in ca 45% of simulation frames with maximum and average observed lifetimes of 203 ns and 68 ns, respectively. An essential observation was that long-term binding of a single methylamine in the m^6A_3 binding pocket on timescale of hundreds of nanoseconds would occur only in situations where the incoming methylamine would successfully displace all water molecules from nearby the W431 side-chain and assume a position very similar to the amino group of m^6A (Supplementary Figure S1). This is consistent with a recent study which showed that release of a water molecule from this region is advantageous in terms of free energy [18]. As expected, the methylamine was not completely selective for the m^6A_3 binding pocket and bound to YTH also at other locations (Supplementary Figure S2) although these were considerably less attractive, with the strongest alternative binding site having occupancy and maximum lifetime of 11% and 73 ns, respectively. Due to the imperfect selectivity of the methylamine, we cautiously refrain from making too strong conclusions from these simulations. However, they provided an important indication that displacement of water molecules from between the incoming nucleotide and the W431 side-chain is a necessary step for the substrate binding. Such result is consistent with the active role of water molecules in binding of RNAs by proteins, as shown by earlier studies [19, 51, 52].

3.6. G_2 nucleotide is specifically recognized by a combination of direct H-bonding and water-mediated interactions.

In the NMR structure, the position and *syn* orientation of G_2 is firmly established by several NOEs [13]. G_2 base stacks with L383 side-chain and forms the G_2(O6)/V385(N) H-bond, both of which were maintained in our simulations except of reversible fluctuations. Simulations further predicted a G_2(N1)/N386(OD1) H-bond (Supplementary Figure S3). All of these interactions are specific for guanosine and are the likely basis of the preferential recognition of 5'-Gm^{6}A-3' motifs by YTHDC1 [13, 22, 24]. However, in addition to the direct H-bonds and base stacking, there was also a water bridge between G_2(N2) and D479(OD) (Figure 5) formed in simulations. It was yet another water-bridge formed between RNA and D479 side-chain, the latter of which also coordinates water molecules involved in m^{6}A$_3$ binding (see above and Figure 2). We subsequently performed simulations where we substituted G_2 with inosine, thus abolishing the water-bridge between G_2 and D479, but keeping the direct protein/RNA interactions via G_2(O6) and G_2(N1) atoms intact (Supplementary Figure S3). These simulations showed that the absence of the water bridges formed by G_2(N2) had broader influence on network of water-mediated interactions in the region, including the water bridges near the m^{6}A$_3$ binding pocket. In presence of I_2, the side-chain shifted its position, seemingly compensating for absence of G_2(N2) amino group and the associated changes in the network of water bridges (Figure 5). The significance of these changes for protein/RNA binding affinity is unclear but it was previously shown by some of us that abolishment of intermolecular as well as intramolecular water bridges can lower the binding affinity of protein/RNA complexes [19]. Therefore, we quantified the effects of the $G_2 \rightarrow I_2$ substitution on RNA binding affinity by performing TI free-energy calculations and NMR and ITC experiments. All three methods explicitly account for changes in the water network. NMR can additionally reveal differences between the G_2 and I_2 protein/RNA binding in terms of chemical shift perturbations (CSPs).

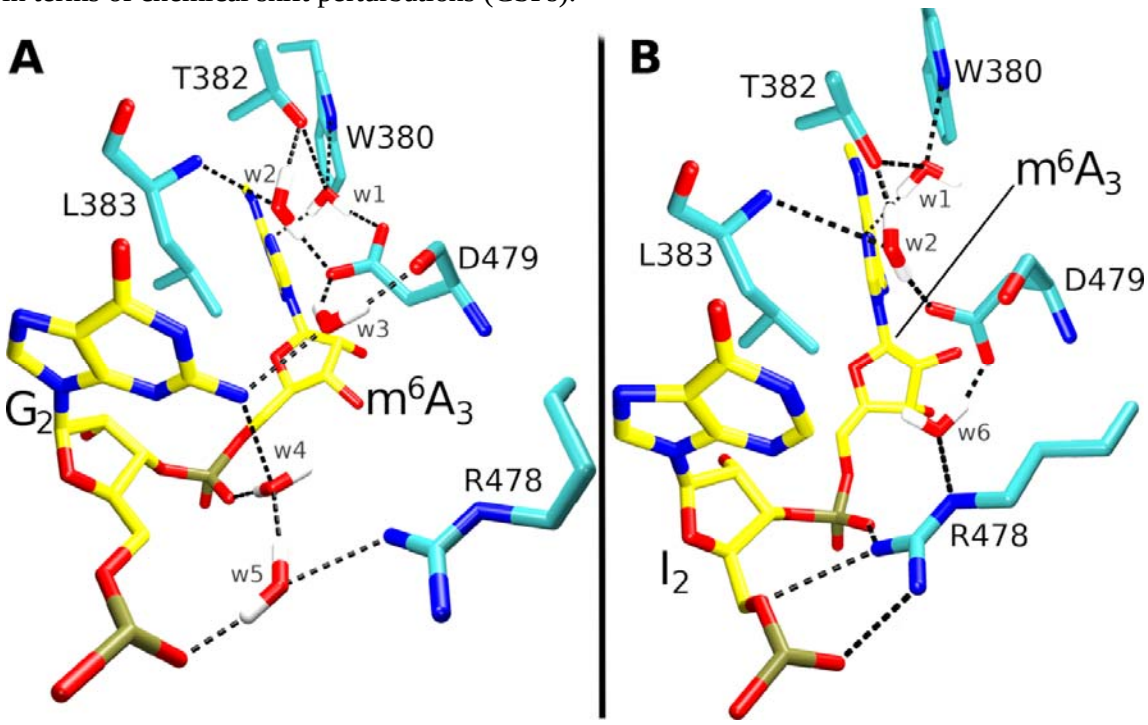

Figure 5. (**A**) Detailed view of the network of water-mediated interactions connecting the G_2 and m^{6}A$_3$ binding pockets in MD simulations. (**B**) Absence of the N2 amino group in I_2 disrupts the arrangement and causes changes to propagate across the entire network. The individual water bridges and protein/RNA interactions are labeled.

3.7. Both calculations and experiments indicate a minor reduction in complex binding affinity for the I_2 RNA.

We performed two independent TI calculations of the $G_2 \rightarrow I_2$ alchemical change that yielded $\Delta\Delta G$ values of 1.0 ± 1.3 and 1.5 ± 0.5 kcal/mol, respectively, for the relative stability of the two complexes. Although the trend is clear, we note the relatively large statistical uncertainty of both calculations, which, in case of the first calculation, exceeds the absolute value of the result. This may be caused by insufficient sampling of the water molecules which constitute the water bridges and directly interact with the atoms of the base undergoing alchemical transformation. For further details of the computational setup and results, see Supplementary.

We also performed ITC measurements to experimentally determine the effect of the $G_2 \rightarrow I_2$ substitution on the complex binding affinity. The ITC yielded K_D values of 25 nM and 40 nM for the G_2- and I_2-containing substrates, respectively (Figure 6). Thus, the ITC also indicated a reduction of binding affinity, albeit merely by ~0.3 kcal/mol.

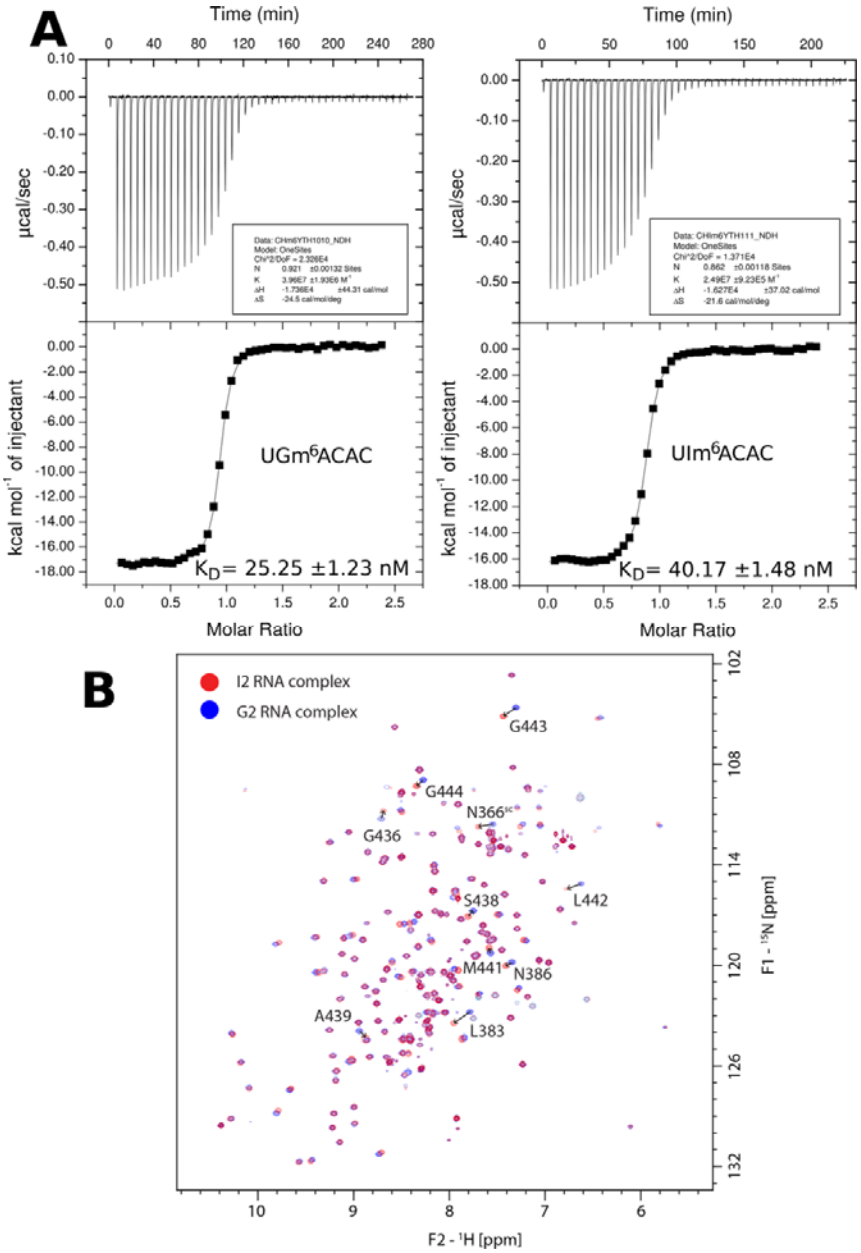

Figure 6. (**A**) ITC data recorded with the YTH domain of YTHDC1 bound to 5′-UGm^{6}ACAC-3′ (left) or 5′-UIm^{6}ACAC-3′ (right) RNAs. The K_D values are shown. Note that the earlier studies show variations in absolute K_Ds for the G_2 substrate which we attribute to different compositions of the experimental buffers [13, 22, 24]. (**B**) Superposition of ^{15}N,^{1}H-HSQC spectra of G_2 and I_2 protein/RNA complexes with 1:1 molecular ratio. Assignments of signals with largest chemical shift changes between the two complexes are shown and their directions indicated by arrows.

3.8. NMR spectroscopy details structural effects of the $G_2 \rightarrow I_2$ substitution.

The ^{15}N,^{1}H-HSQC spectra of the G_2 protein/RNA complex measured for this work were highly similar to those measured for the published NMR structure of the YTH domain of YTHDC1 [13]. We attribute the small differences in peak positions to differences in NMR buffer compositions (see Methods). The I_2 protein/RNA complex, measured for the first time

in this work, also yielded very similar spectra (Figure 6) with only minor differences. Interestingly, the chemical shift changes were observed for the residues involved in the water interaction network between G_2 and m^6A_3 as suggested by the simulations (Figure 5). For example, there was a minor downfield shift of 0.065 ppm for the W380(HE2) atom which is located in the m^6A_3 binding pocket and which is water-bridged with the D479(OD) in G_2 but not in I_2 simulations. A downfield shift of 0.170 ppm was also observed for the L383 backbone amide. According to the simulations, this atom is also involved in the aforementioned water network (Figure 5). This data confirms that there is a weak structural link between the G_2 and m^6A_3 binding pockets; a link which is influenced by the I_2 substitution. Therefore, even though we cannot directly observe the water network with NMR, we suggest the observed chemical shift differences of the specific protein residues are consistent with its existence as suggested by the simulations.

Additional chemical shift differences were observed (Supplementary Figure S4), most of them modest and localized directly around the G_2/I_2 nucleotide binding site. Namely, there was a 0.065 ppm downfield shift of the ^1H signal of N386 backbone amide which was predicted by simulations to interact with the O6 atom of both G_2 and I_2 (Supplementary Figure S3). There was also a strong downfield shift of 0.155 ppm detected for one of the N366 side-chain amide signals (Figure 6). In simulations, this residue sometimes formed an H-bond with the m^6A_3 phosphate which in turn formed an intramolecular water bridge towards the N2 atom of G_2 (Figure 5) as well as a water bridge with the R478 side-chain. In 2mtv_I_2 simulations, these water bridges were abolished in favor of a direct interaction between the m^6A_3 phosphate and R478 (Figure 5). The downfield shift suggests that the strength or population of the H-bond between N366 and the m^6A_3 phosphate was greater in the I_2 system and we suggest this may be the result of electronic structure alteration of the phosphate induced by the different interaction it forms with positively charged R478 side-chain. The remaining significant chemical shift differences were observed for the backbone amides of the β4/β5 protein loop (residues 436 to 446). The experimental structures do not reveal a direct relation between this region and the G_2/I_2 nucleotide binding site. However, we observed transient vdW interactions and H-bonds formed between the loop and the m^6A_3 phosphate in simulations. The L442 is also one of the residues lining the m^6A_3 binding pocket (Figure 2). Lastly, we estimate the contribution of ring current intensities of hypoxanthine and guanine to the chemical shifts differences described above to be negligible [53, 54]. We base this on the small size of ring current shifts calculated for the G_2 complex and the fact that the ring current intensities are only 8% larger for hypoxanthine compared to guanine [55].

3.9. D479A substitution destabilizes binding of the m^6A_3.

In our 2mtv_m^6A_3 simulations, the D479 side-chain formed a centerpiece of the water network connecting the G_2 and m^6A_3 binding pockets and so we decided to perform simulations with YTH D479A mutant (Table 1) which should abolish the water network. As expected, the water bridges directly coordinated by D479 (see Figure 5) were abolished in presence of A479. The influence on the rest of the water molecules, not directly coordinated by D479, was more limited (Table 2). However, there was a curious effect on the w1 water bridge which directly participates in recognition of the m^6A_3 base (Figure 5). Namely, the

water at this position would sometimes fail to properly exchange with the bulk solvent, leaving the site temporarily unoccupied and the many donors and acceptors engaged by this water unsatisfied. The overall effect was that the occupancy of the w1 water bridge was significantly reduced by the D479A substitution (Table 2). In 2mtv_m^6A_3 simulations, the exchange of water molecules forming the w1 water bridge was always seamless, with the departing water molecule immediately replaced by another one from the coordination shell of the negatively charged D479 side-chain. Such mechanism of exchange cannot take place in the 2mtv_D479A_m^6A_3 system and as a result, it could take time for new water to spontaneously enter the pocket after the previous w1 water is expelled in simulations. In the simulation frames where the w1 water bridge was temporarily absent, we sometimes observed reversible shift of the m^6A_3 base into the void left behind by the missing w1 water bridge. This shift was also associated with formation of an alternative "w1a" water bridge between T382(OG) and m^6A_3(O2′) atoms. In 2mtv_m^6A_3 simulations, both of these atoms were forming water-mediated and direct interactions, respectively, with D479 (Figure 2 and Figure 5). The system might be attempting to compensate for the absence D479 side-chain by shifting the position of the m^6A_3 nucleotide. However, the shift also leads to loss of the vdW interaction between the N6-methyl group of m^6A_3 and the highly conserved W431.

In conclusion, simulations with the D479A substitution provide further evidence for the importance of the water network for establishing suitable arrangement of H-bond and vdW interactions ideal for m^6A_3 binding. The D479 side-chain facilitates seamless exchanges with the bulk solvent for the structured water molecules within the pocket while also preventing formation of hydration patterns disruptive for the m^6A_3 binding.

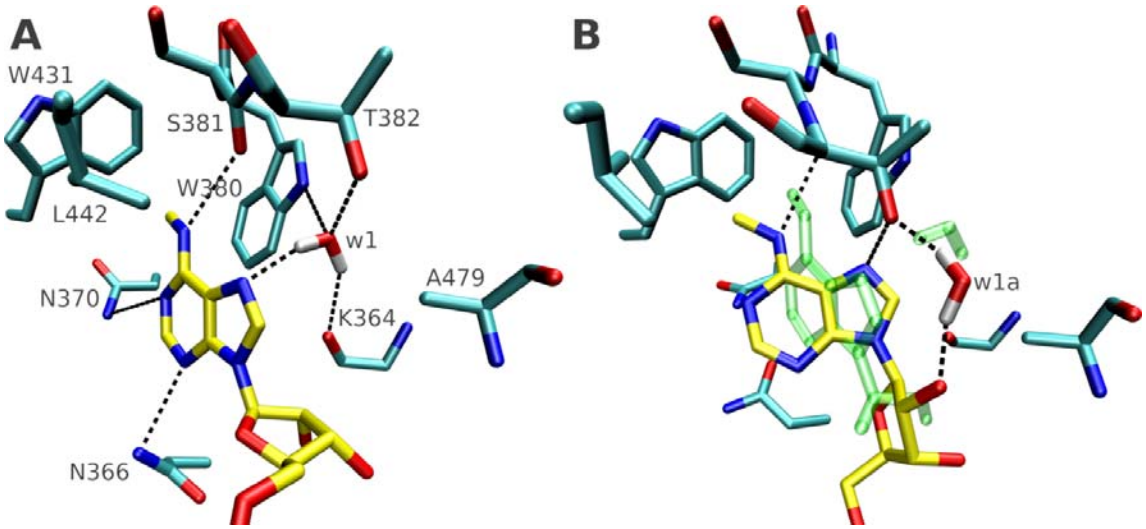

Figure 7. (**A**) Detail of the m^6A_3 binding pocket in 2mtv_D479A_m^6A_3 simulation when w1 water bridge was present. The amino-acids forming the pocket and participating in the water bridge are labeled. (**B**) In some cases, the w1 water failed to exchange with the bulk which sometimes caused the m^6A_3 nucleotide to move into the resulting void and form an alternative w1a water bridge. The shift disrupted many of the protein/RNA interactions native to the m^6A_3 binding pocket. The position of the m^6A_3 nucleotide and w1 water from panel A is shown as green transparent overlay for comparison.

Table 2. Occupancies and maximum observed lifetimes of the structured water molecules in MD simulations of the individual systems.[a]

water molecule / simulation ensemble	occupancies (in %)					
	w1	**w2**	**w3**	**w4**	**w5**	**w6**
2mtv_m^{6}A$_3$	90	52	53	42	50	-
2mtv_m^{6}A$_3$_NOEfix312	90	71	55	55	68	-
2mtv_m^{6}A$_3$_NOEfix18	86	55	57	52	45	-
2mtv_A$_3$[b]	84	46	59	41	52	-
2mtv_I$_2$	97	80	-	-	-	69
2mtv_D479A_m^{6}A$_3$	60	-	-	59	49	-
	lifetimes (in ns)					
	w1	**w2**	**w3**	**w4**	**w5**	**w6**
2mtv_m^{6}A$_3$	660 ± 166	446 ± 253	280 ± 146	37 ± 35	31 ± 46	-
2mtv_m^{6}A$_3$_NOEfix312	624 ± 129	426 ± 69	121 ± 130	29 ± 32	44 ± 50	-
2mtv_m^{6}A$_3$_NOEfix18	604 ± 162	411 ± 122	156 ± 161	26 ± 38	55 ± 37	-
2mtv_A$_3$[b]	363 ± 324	148 ± 65	186 ± 97	33 ± 32	23 ± 23	-
2mtv_I$_2$	703 ± 66	444 ± 35	-	-	-	80 ± 37
2mtv_D479A_m^{6}A$_3$	418 ± 250	-	-	37 ± 21	7 ± 8	-

[a]Designations of the individual water molecules are the same as in Figure 5. The cells marked with "-" indicate that the water bridge was not formed in the specified simulation ensemble. Occupancies indicate percentages of the simulation ensemble in which the specified water bridges were present. An individual water bridge was considered present when its water molecule participated in H-bonding interactions as shown in Figure 5. Lifetimes indicate maximum time a single water molecule was observed forming the individual water bridges as shown in Figure 5, averaged over the individual simulations of the ensemble. Note that in some trajectories, the w1 and w2 water did not exchange on the simulation timescale. [b]Simulation frames where the A$_3$ nucleotide temporarily or permanently drifted out of its binding pocket (see above and Figure 3) were not included in the analyses.

3.10. Use of NOEfix function to further stabilize MD simulations of NMR structures.

A common way to evaluate simulation force-field performance is comparison with the upper-bound NOE distances measured by NMR spectroscopy [38, 56-60]. Here, we observed that over 90% of both intramolecular and intermolecular NOEs were satisfied in the standard 2mtv_m^{6}A$_3$ simulation ensemble (Table 3). We suggest such performance allows making biologically relevant conclusions from simulations that can, furthermore, be experimentally verified (see above and Ref. [38]). Still, some of the NOE distance violations in our simulations occurred in vicinity of the G$_2$ and m^{6}A$_3$ binding pockets (the primary interest of our study; see above). To verify our simulations and further improve the overall agreement with the experimental data, we tested a new approach termed the "NOEfix" potential (Figure 8), which is formally analogous to HBfix, gHBfix and tHBfix approaches tuning stability of H-bonds [40, 49, 61]. NOEfix introduces a mild stabilizing 1 kcal/mol energy bias between two hydrogens within interval of $[d;\ d\ \times 1.3]$ where d is experimental NOE upper bound. Unlike traditional flat-well distance restraints commonly utilized for NMR structure

refinement, the NOEfix does not produce any force potential for distances outside of the specified distance range (Figure 8).

In our initial test, we applied NOEfix for all the violated NOE distances observed in 2mtv_m^{6}A$_3$ ensemble (considering all eight simulations together), corresponding to 312 NOE distances. In this 2mtv_m^{6}A$_3$_NOEfix312 simulation ensemble, we observed significant improvement in the number of both intra- and intermolecular satisfied NOEs in simulations (Table 3). Specifically, 221 NOE violations which were observed in the 2mtv_m^{6}A$_3$ ensemble were corrected in 2mtv_m^{6}A$_3$_NOEfix312 ensemble. Of the remaining violations, 91 were observed in both ensembles and 79 violations were newly observed in the 2mtv_m^{6}A$_3$_NOEfix312. Importantly, the absolute values of the remaining violations were significantly lowered compared to the standard ensemble (Supplementary Figure S5). Still, the appearance of new violations suggests that using 312 NOEfixes is excessive. Also, the thermal fluctuations of some of the residues lining the m^{6}A$_3$ binding pocket were notably reduced in the 2mtv_m^{6}A$_3$_NOEfix312 ensemble (Figure 4) and some of the NOE upper bound distances for which we applied NOEfix were not exceeded at any point of the simulation (Supplementary Figure S6). Likewise, the potential energy associated with NOEfix quickly converged and remained close to constant in these simulations (Supplementary Figure S7). This goes against the original idea of including experimental information in the simulation ensemble without hindering the ability to explore conformational space. In other words, such indiscriminate use of NOEfix has likely a similar effect as mere restraining.

Therefore, in our second test, we limited application of NOEfixes for the violated NOE distances occurring for residues within 6 Å of the bound RNA. Further, only the single largest NOE distance violation associated with each protein residue was considered, obtaining a selection of 18 NOEfixes. See Supplementary for the complete list. In this 2mtv_m^{6}A$_3$_NOEfix18 ensemble, we observed improvement in the intermolecular satisfied NOEs that was on par with the 2mtv_m^{6}A$_3$_NOEfix312 ensemble despite significantly fewer NOEfixes being used. Overall, we observed 64 fewer NOE violations in 2mtv_m^{6}A$_3$_NOEfix18 compared to 2mtv_m^{6}A$_3$, though 35 new NOE violations appeared.

Importantly, use of NOEfix lowered number of NOE distance violations around the G$_2$ and m^{6}A$_3$ binding pockets (Supplementary Figure S8) in both 2mtv_m^{6}A$_3$_NOEfix312 and 2mtv_m^{6}A$_3$_NOEfix18 simulation ensembles. Almost the same improvement could be achieved with 18 NOEfixes and the use of fewer, carefully selected NOEfixes does not repress the dynamics (Figure 4 and Supplementary Figures S6 and S9). It should be noted that ultimately, for the present system, the NOEfix ensembles provided similar structural information as the 2mtv_m^{6}A$_3$ ensemble (Figure 4 and Figure 5). Thus the main results reported in this work are not affected by force-field related deviations of the simulations with respect to the primary experimental data. This is because the RNA binding to YTH is sufficiently well determined by the experimental data and well described by the force field. Possible uses and advantages of NOEfix approach in less well-behaved systems are stated in the Discussion.

Table 3. Percentage of satisfied NOE upper bound distances in MD simulations of the YTH domain complexed with 5′-C$_1$G$_2$m^{6}A$_3$C$_4$A$_5$C$_6$-3′ RNA.

simulated system	percentage of satisfied NOE distances		
	protein	RNA	protein/RNA
2mtv_m^{6}A$_3$	93%	96%	94%
2mtv_m^{6}A$_3$_NOEfix312	96%	98%	99%
2mtv_m^{6}A$_3$_NOEfix18	94%	98%	99%

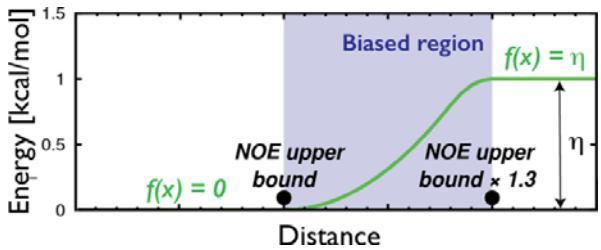

Figure 8. The NOEfix potential (green curve) produces non-zero forces only in a narrow region between the NOE upper bound distance and the same distance multiplied by factor of 1.3. A smoothly increasing potential is applied in this range until it reaches the maximum energy penalty η. In this fashion, a steeper NOEfix potential is utilized for stronger (shorter) NOEs while a more gentle potential is present for the weaker (longer) NOEs. The distance range and η values are adjustable. For details of the technical implementation of the potential function, see Refs. [40, 49, 61]. NOEfix is a pragmatic means of introducing some bias based on the primary NMR data to compensate for the force-field deficiencies and to verify the performance of fully unbiased simulations while maintaining a wider range of conformations than what would be observed with traditional restraints where the energy penalty keeps increasing with distance.

4. Discussion

4.1. Structural hydration plays a key role in recognition of the N6-methylated adenosine by the YTH domain and in discrimination against unmethylated adenosine.

Our simulations provided unique insights into interrelation between hydration dynamics and protein/RNA recognition by revealing a set of water bridges in the vicinity of the m^{6}A$_3$ binding pocket. A key role of water molecules for the m^{6}A$_3$ binding and discrimination against A$_3$ was further shown by our simulations with the unmethylated A$_3$ nucleotide. In the bound state, the A$_3$ nucleotide had two dynamically competing binding states which both resembled the ground state of m^{6}A$_3$ when considering the position of the base. In the dominant state (~90% occupancy), there was no water molecule inside the pocket. This left an unsaturated H-bond donor as well as not fully saturated vdW surface within the pocket due to absence of the N6-methyl group on A$_3$. In other words, in the dominant binding state, the protein could not fully optimize its interactions with the unmethylated adenosine. In our simulations, the system consequently sampled an additional auxiliary binding state (~9% of time) where a single water molecule transiently invaded the binding pocket in the A$_3$ simulations (Figure 3). This water molecule occupied the space in which the N6-methyl group would normally be located for the m^{6}A$_3$ substrate, forming a single H-bond with the H61 hydrogen of A$_3$, as well as compensating for the loss of the methyl-group vdW contacts

(Figure 3). However, this single water molecule was unable to establish any other H-bonding interactions. Such an arrangement likely carries a large free-energy penalty since a single water molecule can potentially form up to four H-bonds [62]. Furthermore, buried water molecules are entropically unfavorable [63]. Thus, for distinct structural reasons, the A_3 molecular interactions with YTH were suboptimal both in the dominant and auxiliary binding states.

What is striking is that the auxiliary binding state appears to also be a transition (or intermediate) state along the coordinate corresponding to the (un)binding process for A_3. Specifically, our simulations showed that the invading water molecule associated with the auxiliary state was either expelled again from the pocket after some time, thus reverting to the dominant binding state, or it attracted additional water molecules from the bulk which, then, gradually displaced the A_3 nucleotide from the pocket (Figure 3). Based on this data, we suggest that binding of m^6A_3 and A_3 substrates by the YTH domain differs not only in the free energy of the final states, as shown by the experiments [13, 22], but that the kinetics and pathways of the two bindings could also be different. Namely, the recognition of the m^6A_3 and A_3 by the YTH could correspond to two-state and four-state bindings, respectively (Figure 9). This could have biological implications as the dynamics of the single water molecule observed in the auxiliary binding state of unmethylated A_3 and its ability to attract even more waters from the bulk could aid in more frequent displacements of the unmethylated adenosine from the pocket. In the presence of other RNA molecules, it could lead to a faster substrate cycling, and, in turn, allow the YTH domain to more quickly locate the N6-methylated adenosine among a pool of cellular RNAs where unmethylated adenosine can be expected to be much more prevalent. Incompatibility between the A_3 binding and relaxed water network is assumed to substantially increase "local" k_{off} kinetic constant of the nucleotide.

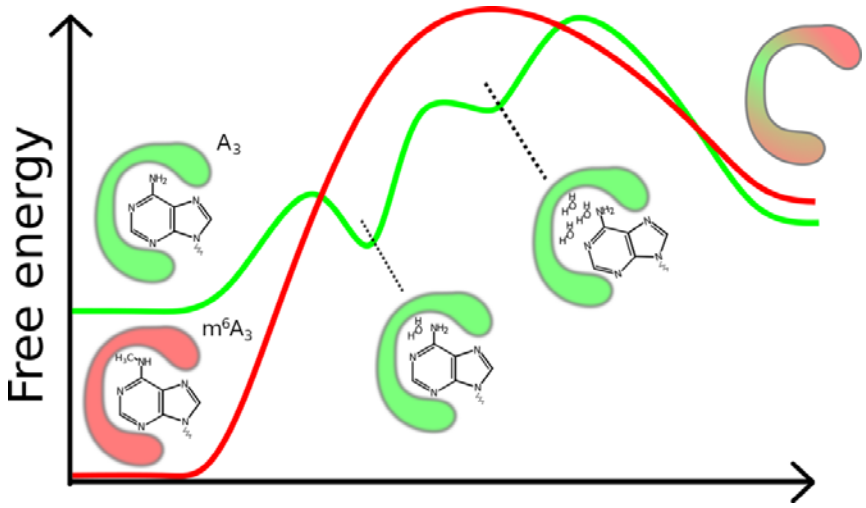

Figure 9. Scheme of the differences in binding states of the N6-methylated (red) and unmethylated (green) adenosine by the YTH domain of YTHDC1. The energy curves are merely illustrative of simulation trends and are not a result of measurements or free-energy computations.

Lastly, we note that our simulations of the free YTH domain showed that the binding pocket is preformed and fully accessible to the solvent; in other words, it is hydrated at all times (Figure 3). Thus, the substrate must displace these water molecules during the binding. The A_3 Watson-Crick base edge is presumably more attractive for waters and could be less effective in this process than m^6A_3 due to the hydrophobic methyl group. The importance of displacing the water molecules from the pocket was shown by simulations of the free YTH domain with methylamine molecules (a molecular mimic of the N6-methylated amino group) which were able to (meta)stably bind to the YTH domain binding pocket only after excluding all water molecules from the pocket (Supplementary Figures S1 and S2). Position of the bound methylamine confirms that the pocket is perfectly optimized to recognize the modified nucleotide. We acknowledge the simulations are likely not long enough to capture all the possible structural micro-arrangements. Nevertheless, they provide an illuminating view into the role of water dynamics during binding/unbinding of a N6-methylated adenosine in the binding pocket of YTH and how this can be used for discrimination against an unmethylated substrate. We propose that similar mechanisms might be used by some other proteins which recognize residue methylation through similarly structured binding pockets [8, 11].

4.2. The YTH domain can tolerate inosine as the nucleotide preceding m^6A_3.

Earlier experiments have shown that the YTHDC1 binds the RNA substrate more strongly when guanosine is the nucleotide immediately preceding the m^6A (the G_2 in our work) [13, 15, 22]. Our simulations showed that the G_2 formed several specific albeit fluctuating interactions with the YTH. Among these was a water bridge interaction between the N2 amino group and the D479 side-chain (Figure 5). We performed several simulations in which we mutated the G_2 into inosine to abolish this water bridge. We subsequently quantified its effect on RNA binding affinity by performing TI calculations and ITC measurements (Figure 6A), which indicated a minor free-energy penalty for the I_2 complex of 1.0-1.5 kcal/mol and 0.3 kcal/mol, respectively. Furthermore, NMR spectroscopy identified residues affected by this modification and suggested that the absence of the N2 amino group in the I_2 RNA complex and the loss of the water bridge are partially compensated for by strengthening of some of the other protein/RNA interactions. Our work thus suggests that the YTH domain of YTHDC1 can recognize both 5′-Gm^6A-3′ and 5′-Im^6A-3′ motifs with only a minor free-energy penalty for the latter. Im^6A-rich RNA sequences could also be potentially recognized by YTH, even though it was proposed that adenosine deamination (leading to inosine) and N6-methylation are mutually exclusive chemical modifications within a single transcript [64].

4.3. Water bridges facilitate communication between the G_2 and m^6A_3 binding pockets.

The simulations revealed a network of water bridges, connecting the G_2 nucleotide through its 2-amino group with residues lining the m^6A_3 binding pocket (Figure 5). Lifetimes of the individual waters in this network were on timescale of tens to hundreds of nanoseconds (Table 2) while common hydrations sites around biomolecules have sub-ns time scale [16]. We showed that the water network could be altered by $G_2 \rightarrow I_2$ substitution which abolishes water bridges made by the 2-amino group (Figure 5), although the direct H-bonds with the protein are not affected. Chemical shift changes were detected by NMR for the protein

residues deep within the m^6A_3 binding pocket, giving an indirect experimental support for existence of the water network (see above and Figure 6B). We further explored the intricacies of the water network in our simulations of the YTH protein/RNA complex with D479A mutation. According to simulations, the side-chain of D479 is specifically involved in coordinating the water network (Figure 5) and its removal had profound effects. Namely, without D479, the single water bridge deep in the m^6A_3 binding pocket was sometimes unable to make a complete exchange of water with the bulk solvent, leaving temporarily to a vacant hydration site. This likely would compromise the binding affinity of m^6A_3 (Figure 7). This illustrates the importance of the water molecules both for establishing the specificity of the N6-methylation recognition (see above and Figure 3) as well as proper arrangement of the binding pocket. In other words, hydration is an important element of the auxiliary sequence-specificity for G_2 that YTHDC1 evolved as an extension of the m^6A_3 binding site.

4.4. Agreement between MD simulations and experimental data can be improved by application of NOEfix corrections.

The simulations showed a reasonable force-field performance in describing parts of the protein/RNA interface related to recognition of the G_2 and m^6A_3 nucleotides (the primary focus of this study). Nevertheless, there were some NOE distance violations reported for this region in the standard simulations. We did not wish to restrain these distances by traditional NMR restraints as doing so would have severely limited the dynamics. Therefore, we decided to instead apply the newly-developed "NOEfix" function (Figure 8). Based on the same underlying math as the HBfix [16, 40], the NOEfix increases the agreement between the simulation and the experimentally measured upper-bound NOE distances (Table 3). In the two simulation tests, we applied the NOEfix to either all observed NOE distance violations (312 NOEfixes), or to selected NOE distance violations occurring close to the protein/RNA interface (18 NOEfixes). We conclude that latter choice is more appropriate as using too many NOEfixes at the same time can repress the structural dynamics (Figure 4 and Supplementary Figure S6). In both tests, the use of NOEfix corrected majority of the violated NOE distances but also introduced some new violations. This shows that simultaneous satisfactory description of all the NOE distances indicated by the experiment is not favored by the force field. While not fully successful, it is important to note that a perfect agreement with NMR data should not be realistically expected due to inherent errors of the experimental technique as well as the unphysical approximations of the molecular mechanical description. The NMR data used for comparison is also ensemble- and time-averaged [21] and so the inability to satisfy all the NOE distances in simulations could also reflect insufficient sampling. The appearance of new NOE violations could also be a sign of an already excessive bias even when using only 18 NOEfixes. Possible improvements could be achieved by reducing the number of utilized NOEfixes even further. The system studied in this work is also well-determined, possessing very high density of NMR signals, many of which are structurally coupled [13]. The NOEfix could be of greater benefit in systems where insufficient number of signals was detected due to, e.g. partial disorder or conformational transitions.

Obviously, the NOEfix approach does not represent a direct rigorous integration of the primary NMR data into the MD simulation, such as for example the maximum entropy approaches [65, 66]. However, it is easy and flexible in its implementation and can improve quality of the simulated ensembles even in cases when direct integration of the experimental data into simulation protocols is inconvenient due to, e.g., insufficient amount of the primary experimental data or large size of the system. NOEfix could be particularly useful for resolving distance violations arising from minor force-field inaccuracies across multiple terms whose direct tuning would lead to problems in other parts of the system. The presently suggested NOEfix parameters were chosen rather intuitively and are not based on any rigorous approach. We plan to test the method in the near future for a wide range of protein-RNA complexes, including those which are less well determined and show larger instabilities in the simulations.

4.5. Concluding remarks.

We have investigated dynamics of the YTH domain of the YTHDC1 protein and its binding to N6-methyladenosine containing RNA. We show that the mechanism of recognition, in addition to stabilization via the vdW interactions with the methyl group, includes contributions from structured water molecules. While long-residency water molecules are helping to cement binding of the adenine moiety, unmodified adenosine is unable to keep water molecules from within the N6-methyladenosine binding pocket. Our study therefore provides an additional structural basis for the ~50× reduction in affinity of unmodified adenosine containing RNAs compared to those harboring m^6A. Water molecules are also involved in recognition of the G_2 residue and there is a network of water bridges connecting the two binding pockets. We used NMR to measure the effects of disrupting this network. The binding of N6-methyladenosine is also not stable when the water network is abolished because of a single amino-acid mutation. Lastly, we describe a method for introducing a weak bias towards the primary NMR NOE data which could help to improve simulations of some protein-RNA complexes. In this manner the simulations starting from the bound state can be stabilized while avoiding excessive suppression of the structural dynamics.

Supporting Information

Supporting Information are available online.

Acknowledgement

This work was supported by the Czech Science Foundation [grant number 20-16554S to M.K. and J.S]; NCCR RNA and Disease from the Swiss National Science Foundation to F.A.; and CESNET storage facilities [grant number CESNET LM2018140].

References

[1] W.A. Cantara, P.F. Crain, J. Rozenski, J.A. McCloskey, K.A. Harris, X. Zhang, F.A.P. Vendeix, D. Fabris, P.F. Agris, The RNA Modification Database, RNAMDB: 2011 Update, Nucleic Acids Res., 39 (2011) D195-D201.
[2] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, Topology of the Human and Mouse m6A RNA Methylomes Revealed by m6A-seq, Nature, 485 (2012) 201-206.
[3] X. Wang, Z. Lu, A. Gomez, G.C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan, C. He, N6-methyladenosine-dependent Regulation of Messenger RNA Stability, Nature, 505 (2013) 117.
[4] X. Wang, Boxuan S. Zhao, Ian A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, N6-methyladenosine Modulates Messenger RNA Translation Efficiency, Cell, 161 (2015) 1388-1399.
[5] Kate D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, Christopher E. Mason, Samie R. Jaffrey, Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons, Cell, 149 (2012) 1635-1646.
[6] B. Wu, L. Li, Y. Huang, J. Ma, J. Min, Readers, Writers and Erasers of N6-methylated Adenosine Modification, Curr. Opin. Struct. Biol., 47 (2017) 67-76.
[7] M. Reichel, T. Köster, D. Staiger, Marking RNA: m6A Writers, Readers, and Functions in Arabidopsis, Journal of Molecular Cell Biology, 11 (2019) 899-910.
[8] S. Zaccara, R.J. Ries, S.R. Jaffrey, Reading, Writing and Erasing mRNA Methylation, Nat. Rev. Mol. Cell Biol., 20 (2019) 608-624.
[9] Z. Zhang, D. Theler, K.H. Kaminska, M. Hiller, P. de la Grange, R. Pudimat, I. Rafalska, B. Heinrich, J.M. Bujnicki, F.H.-T. Allain, S. Stamm, The YTH Domain Is a Novel RNA Binding Domain, J. Biol. Chem., 285 (2010) 14701-14710.
[10] S. Luo, L. Tong, Molecular Basis for the Recognition of Methylated Adenines in RNA by the Eukaryotic YTH Domain, Proc. Natl. Acad. Sci. U.S.A., 111 (2014) 13834-13839.
[11] S. Khorasanizadeh, Recognition of Methylated Histones: New Twists and Variations, Curr. Opin. Struct. Biol., 21 (2011) 744-749.
[12] A.M. Hartmann, O. Nayler, F.W. Schwaiger, A. Obermeier, S. Stamm, The Interaction and Colocalization of Sam68 with the Splicing-associated Factor YT521-B in Nuclear Dots Is Regulated by the Src Family Kinase p59fyn, Mol. Biol. Cell, 10 (1999) 3909-3926.
[13] D. Theler, C. Dominguez, M. Blatter, J. Boudet, F.H.T. Allain, Solution Structure of the YTH domain in Complex with N6-methyladenosine RNA: A Reader of Methylated RNA, Nucleic Acids Res., 42 (2014) 13911-13919.
[14] Y. Li, R.K. Bedi, L. Wiedmer, D. Huang, P. Śledź, A. Caflisch, Flexible Binding of m6A Reader Protein YTHDC1 to Its Preferred RNA Motif, J. Chem. Theory Comput., 15 (2019) 7004-7014.
[15] A.E. Arguello, R.W. Leach, R.E. Kleiner, In Vitro Selection with a Site-Specifically Modified RNA Library Reveals the Binding Preferences of N6-Methyladenosine Reader Proteins, Biochemistry, 58 (2019) 3386-3395.
[16] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R.A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N.G. Walter, M. Otyepka, RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview, Chem. Rev., 118 (2018) 4177–4338.
[17] Y. Li, R.K. Bedi, E.V. Moroz-Omori, A. Caflisch, Structural and Dynamic Insights into Redundant Function of YTHDF Proteins, J. Chem. Inf. Model., (2020).

[18] Y. Li, R.K. Bedi, L. Wiedmer, X. Sun, D. Huang, A. Caflisch, Atomistic and Thermodynamic Analysis of N6-Methyladenosine (m6A) Recognition by the Reader Domain of YTHDC1, J. Chem. Theory Comput., (2021).
[19] M. Krepl, M. Blatter, A. Cléry, F.F. Damberger, F.H.T. Allain, J. Sponer, Structural Study of the Fox-1 RRM Protein Hydration Reveals a Role for Key Water Molecules in RRM-RNA Recognition Nucleic Acids Res., 45 (2017) 8046-8063.
[20] S. Khodadadi, A.P. Sokolov, Atomistic Details of Protein Dynamics and the Role of Hydration Water, Biochimica et Biophysica Acta (BBA) - General Subjects, 1861 (2017) 3546-3552.
[21] S. Campagne, M. Krepl, J. Sponer, F.H.T. Allain, Chapter Fourteen - Combining NMR Spectroscopy and Molecular Dynamic Simulations to Solve and Analyze the Structure of Protein–RNA Complexes, in: A.J. Wand (Ed.) Methods Enzymol., vol. 614, Academic Press, 2019, pp. 393-422.
[22] C. Xu, X. Wang, K. Liu, I.A. Roundtree, W. Tempel, Y. Li, Z. Lu, C. He, J. Min, Structural Basis for Selective Binding of m6A RNA by the YTHDC1 YTH Domain, Nat. Chem. Biol., 10 (2014) 927-929.
[23] C.B. Woodcock, J.R. Horton, J. Zhou, M.T. Bedford, R.M. Blumenthal, X. Zhang, X. Cheng, Biochemical and Structural Basis for YTH Domain of Human YTHDC1 Binding to Methylated Adenine in DNA, Nucleic Acids Res., in press (2020) 10.1093/nar/gkaa1604.
[24] C. Xu, K. Liu, H. Ahmed, P. Loppnau, M. Schapira, J. Min, Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins, J. Biol. Chem., 290 (2015) 24902-24913.
[25] I.Y.B.-S. D.A. Case, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 18, (2018).
[26] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. Hauser, C. Simmerling, ff14SB: Improving The Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J. Chem. Theory Comput., 11 (2015) 3696-3713.
[27] A. Perez, I. Marchan, D. Svozil, J. Sponer, T.E. Cheatham, C.A. Laughton, M. Orozco, Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers, Biophys. J., 92 (2007) 3817-3829.
[28] M. Zgarbova, M. Otyepka, J. Sponer, A. Mladek, P. Banas, T.E. Cheatham, P. Jurecka, Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles, J. Chem. Theory Comput., 7 (2011) 2886-2902.
[29] M.J.T. Frisch, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al., Gaussian 09, Revision D01, in, Gaussian, Inc., Wallingford CT, 2009.
[30] P. Cieplak, W.D. Cornell, C. Bayly, P.A. Kollman, Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins, J. Comput. Chem., 16 (1995) 1357-1377.
[31] M. Krepl, M. Otyepka, P. Banas, J. Sponer, Effect of Guanine to Inosine Substitution on Stability of Canonical DNA and RNA Duplexes: Molecular Dynamics Thermodynamics Integration Study, J. Phys. Chem. B, 117 (2013) 1872-1879.

[32] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The Missing Term in Effective Pair Potentials, J. Phys. Chem., 91 (1987) 6269-6271.
[33] S. Izadi, R. Anandakrishnan, A.V. Onufriev, Building Water Models: A Different Approach, J. Phys. Chem. Lett., 5 (2014) 3863-3871.
[34] T. Steinbrecher, J. Latzer, D.A. Case, Revised AMBER Parameters for Bioorganic Phosphates, J. Chem. Theory Comput., 8 (2012) 4405-4412.
[35] V. Mlynsky, P. Kuhrova, M. Zgarbova, P. Jurecka, N.G. Walter, M. Otyepka, J. Sponer, P. Banas, Reactive Conformation of the Active Site in the Hairpin Ribozyme Achieved by Molecular Dynamics Simulations with epsilon/zeta Force Field Reparametrizations, J. Phys. Chem. B, 119 (2015) 4220-4229.
[36] I.S. Joung, T.E. Cheatham, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations, J. Phys. Chem. B, 112 (2008) 9020-9041.
[37] M. Krepl, J. Vögele, H. Kruse, E. Duchardt-Ferner, J. Wöhnert, J. Sponer, An Intricate Balance of Hydrogen Bonding, Ion Atmosphere and Dynamics Facilitates a Seamless Uracil to Cytosine Substitution in the U-turn of the Neomycin-sensing Riboswitch, Nucleic Acids Res., 46 (2018) 6528–6543.
[38] M. Krepl, A. Cléry, M. Blatter, F.H.T. Allain, J. Sponer, Synergy between NMR Measurements and MD Simulations of Protein/RNA Complexes: Application to the RRMs, the Most Common RNA Recognition Motifs, Nucleic Acids Res., 44 (2016) 6452-6470.
[39] S. Le Grand, A.W. Götz, R.C. Walker, SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations, Comput. Phys. Commun., 184 (2013) 374-380.
[40] P. Kuhrova, R. Best, S. Bottaro, G. Bussi, J. Sponer, M. Otyepka, P. Banas, Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies, J. Chem. Theory Comput., 12 (2016) 4534–4548.
[41] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes, J. Comput. Phys., 23 (1977) 327-341.
[42] C.W. Hopkins, S. Le Grand, R.C. Walker, A.E. Roitberg, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, J. Chem. Theory Comput., 11 (2015) 1864-1874.
[43] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large Systems, J. Chem. Phys., 98 (1993) e10089.
[44] T. Steinbrecher, D.L. Mobley, D.A. Case, Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations, J. Chem. Phys., 127 (2007) 214108.
[45] R. Keller, The Computer Aided Resonance Assignment, CANTINA Verlag, Switzerland, 2004.
[46] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, J. Chem. Theory Comput., 9 (2013) 3084-3095.
[47] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual Molecular Dynamics, J. Mol. Graph., 14 (1996) 33-38.
[48] E.A. Merritt, D.J. Bacon, Raster3D: Photorealistic Molecular Graphics, in: C.W. Carter, R.M. Sweet (Eds.) Macromolecular Crystallography, Pt B, vol. 277, Elsevier Academic Press Inc, San Diego, 1997, pp. 505-524.
[49] P. Kuhrova, V. Mlynsky, M. Zgarbova, M. Krepl, G. Bussi, R.B. Best, M. Otyepka, J. Sponer, P. Banas, Improving the Performance of the RNA Amber Force Field by Tuning the Hydrogen-Bonding Interactions, J. Chem. Theory Comput., 15 (2019) 3288-3305.
[50] V. Mlýnský, G. Bussi, Exploring RNA Structure and Dynamics through Enhanced Sampling Simulations, Curr. Opin. Struct. Biol., 49 (2018) 63-71.

[51] A. Barik, C. Nithin, Naga Bhushana R. Karampudi, S. Mukherjee, R.P. Bahadur, Probing Binding Hot Spots at Protein–RNA Recognition Sites, Nucleic Acids Res., 44 (2016) e9.
[52] A. Barik, R.P. Bahadur, Hydration of Protein–RNA Recognition Sites, Nucleic Acids Res., 42 (2014) 10148-10160.
[53] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: A Program for Display and Analysis of Macromolecular Structures, J. Mol. Graph., 14 (1996) 51-55.
[54] D.A. Case, Calibration of Ring-current Effects in Proteins and Nucleic Acids, J. Biomol. NMR, 6 (1995) 341-346.
[55] F.R. Prado, C. Giessner-Prettre, Parameters for the Calculation of the Ring Current and Atomic Magnetic Anisotropy Contributions to Magnetic Shielding Constants: Nucleic Acid Bases and Intercalating Agents, Comput. Theor. Chem., 76 (1981) 81-92.
[56] C. Bergonzo, T.E. Cheatham, Improved Force Field Parameters Lead to a Better Description of RNA Structure, J. Chem. Theory Comput., 11 (2015) 3969-3972.
[57] A. Spasic, S.D. Kennedy, L. Needham, M. Manoharan, R. Kierzek, D.H. Turner, D.H. Mathews, Molecular Dynamics Correctly Models the Unusual Major Conformation of the GAGU RNA Internal Loop and with NMR Reveals an Unusual Minor Conformation, RNA, 24 (2018) 656-672.
[58] S. Bottaro, G. Bussi, S.D. Kennedy, D.H. Turner, K. Lindorff-Larsen, Conformational Ensembles of RNA Oligonucleotides from Integrating NMR and Molecular Simulations, Sci. Adv., 4 (2018) eaar8521.
[59] B. Zagrovic, W.F. Van Gunsteren, Comparing Atomistic Simulation Data with the NMR Experiment: How Much Can NOEs Actually Tell Us?, Proteins: Struct., Funct., Bioinf., 63 (2006) 210-218.
[60] Z. Lin, C. Oostenbrink, W.F. van Gunsteren, On the Use of One-step Perturbation to Investigate the Dependence of NOE-derived Atom–atom Distance Bound Violations of Peptides upon a Variation of Force-field Parameters, Eur. Biophys. J., 43 (2014) 113-119.
[61] V. Mlýnský, P. Kührová, T. Kühr, M. Otyepka, G. Bussi, P. Banáš, J. Šponer, Fine-Tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides, J. Chem. Theory Comput., 16 (2020) 3936-3946.
[62] P. Ball, Water is an Active Matrix of Life for Cell and Molecular Biology, Proc. Natl. Acad. Sci. U.S.A., 114 (2017) 13327-13335.
[63] H. Yu, S.W. Rick, Free Energy, Entropy, and Enthalpy of a Water Molecule in Various Protein Environments, J. Phys. Chem. B, 114 (2010) 11552-11560.
[64] J.-F. Xiang, Q. Yang, C.-X. Liu, M. Wu, L.-L. Chen, L. Yang, N6-Methyladenosines Modulate A-to-I RNA Editing, Mol. Cell, 69 (2018) 126-135.e126.
[65] J.W. Pitera, J.D. Chodera, On the Use of Experimental Observations to Bias Simulated Ensembles, J. Chem. Theory Comput., 8 (2012) 3445-3451.
[66] M. Bonomi, G.T. Heller, C. Camilloni, M. Vendruscolo, Principles of Protein Structural Ensemble Determination, Curr. Opin. Struct. Biol., 42 (2017) 106-116.