Structure-based design of ligands of the m6A-RNA reader YTHDC1

Yaozong Li\textsuperscript{a,b}, Rajiv Kumar Bedi\textsuperscript{a}, Francesco Nai\textsuperscript{a}, Valentin von Roten\textsuperscript{a}, Aymeric Dolbois\textsuperscript{a}, František Zálešák\textsuperscript{a}, Raed Nachawati\textsuperscript{a}, Danzhi Huang\textsuperscript{a}, Amedeo Caflisch\textsuperscript{a,*}

\textsuperscript{a} Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057, Zurich, Switzerland
\textsuperscript{b} Department of Chemistry, Umeå University, SE-901 87, Umeå, Sweden

1. Introduction

Recent studies on post-transcriptional and co-transcriptional modifications at the RNA level have led to a new subject called RNA epigenetics \cite{1} or epitranscriptomics \cite{2}. The N6-methyladenosine (m6A) modification is the most abundant internal chemical modification in messenger RNA (mRNA), and it has been the subject of intense investigations. The m6A deposition is considered a reversible and dynamic process \cite{9}, although these features might require further validation \cite{10}. In a normal functional state, m6A methylation is catalyzed by the METTL3/METTL14 methyltransferase \cite{11,12}, and is erased by the demethylases FTO \cite{13} and ALKBH5 \cite{14}. The downstream effects of m6A modification are elicited by m6A binding proteins (readers). Five direct reader proteins have been discovered so far: YTHDC1 \cite{15}, YTHDC2 \cite{16} and YTHDF1-3 \cite{15,17}. The YTH domain of YTHDC1 contains an aromatic cage evolved specifically for the binding of m6A \cite{27}. Residues surrounding the cage (Fig. 1), protein dynamics, and structural water molecules play essential roles in the binding \cite{28,29}. We have recently investigated the atomistic details of how the YTH domain of YTHDC1 recognizes m6A-containing RNA by combining molecular dynamics (MD) simulations and crystallography \cite{28,30}. We also employed high-throughput docking to identify small-molecule binders. Protein X-ray crystallography and a biochemical assay were used to validate the docking results \cite{31}. As these ligands were in the medium-high micromolar range of affinity, we decided to continue our efforts to search for more potent binders. In the present study, we first used high-throughput docking to screen small-molecule ligands of the m6A-recognition domain of YTHDC1. By combining structural features of two small-molecule ligands, we have designed a low micromolar binder of YTHDC1 which we have validated by X-ray crystallography, a biochemical assay, and isothermal titration calorimetry (ITC).

We report new chemical entities for disrupting the interactions between N6-methyladenosine (m6A) mRNA and its reader YT521-B homology-domain-containing protein 1 (YTHDC1). High-throughput docking was used to screen commercially available databases of small molecules, and molecular dynamics simulations were employed to evaluate the binding stability of m6A nucleotide analogues. The poses of 25 fragment-like new binders were confirmed by X-ray crystallography. The structure-based merging of two weak fragments resulted in a ligand-efficient binder (compound 6) which shows an equilibrium dissociation constant of 1.7 μM in isothermal titration calorimetry measurements and a ligand efficiency value of 0.66 kcal mol\textsuperscript{-1} nHA\textsuperscript{m}.
2. Results and discussion

This work reports 25 small-molecule binders of the m6A-reader domain of YTHDC1. (In the following text, the term YTHDC1 is used frequently as an abbreviation for the m6A-recognition domain of YTHDC1). The new ligands are divided into four groups according to their chemical structures and/or interactions with YTHDC1 (Table 1 and Fig. 2). Group 1 includes analogues of the m6A nucleobase. They were expected to improve the binding affinity and ligand efficiency (which is the binding free energy divided by the number of heavy, i.e., non-hydrogen atoms, \( n_{\text{HA}} \)) by exploring additional interactions with the binding site (Fig. 2a). Group 2 binders form two hydrogen bonds between their N6-methyl amide moiety and two polar residues in the binding pocket (Fig. 2b). The third group of compounds are characterized by interactions between their morpholine (pyrrolidine in compound 24) and the aromatic cage (Fig. 2c). Most of these compounds displace the recognition loop, which results in a larger aperture of the binding site. Group 4 includes a compound that displaces the structural water molecule [28] (Fig. 2d) and other chemotypes (Fig. 2e).

2.1. Improving ligand efficiency by replacing the ribose of the m6A nucleoside

The m6A nucleoside is recognized by the aromatic cage of YTHDC1 in which three residues, namely, Ser378, Asn367, and Asn363, form hydrogen bonds with the nucleobase (Fig. 1). The N6-methyl group further strengthens the ligand-protein binding by van der Waals (vdW) interactions with the lipophilic binding site. Although the ribose ring presents multiple hydrogen-bond interactions with its surrounding residues, these residues are solvent-exposed and thus may not significantly contribute to the binding affinity (Figure S1). We, therefore, first tested the binding affinity of the m6A nucleobase by a homogenous time-resolved fluorescence-based (HTRF) assay [33]. The removal of the ribose group results in a twofold reduction in binding affinity with an improvement in ligand efficiency from 0.23 to 0.38 kcal mol\(^{-1}\) \( n_{\text{HA}} \) (Table 1). This result encouraged us to explore fragment analogues of the m6A nucleobase. We searched the analogues via the small-molecule catalog Chemspace (https://chem-space.com/) and identified ten fragments. We then decided to run multiple MD simulations (of 5 ns each) to evaluate the stability of their predicted binding modes with YTHDC1. We finally decided to test in vitro the five fragments that showed stable binding modes in the MD simulations. One of the five derivatives of the m6A nucleobase was purchased from a commercial vendor (compound 1), while the derivatives 2 to 5 were synthesized in-house. The HTRF assay shows that the fragments 1–3 have a nearly threefold affinity improvement and similar ligand efficiency as the m6A nucleobase (Table 1). Furthermore, the crystal structures of the complex with the YTH domain of YTHDC1 show essentially the same binding mode as the nucleobase of the m6A nucleoside (Figure S2a).

Near the m6A binding site, a positively charged shallow pocket binds the negatively charged RNA (Fig. 1) [30]. Considering the beneficial effect that simple substitutions at N9 had on affinity (methyl in compound 2 and cyclopropyl in compound 3, Figure S1 and S2a), we propose to grow the fragments into the positively charged sub-pocket by linking the nucleobase with polar groups. Another possible position for substitution is C2 between N1 and N3 of the m6A nucleobase (Fig. 2a). The chloride substituent (compound 1) in this position fits a small lipophilic pocket (PDB ID: 7P88). We also tried different substituents in this position, such as the methylamine group (compounds 4 and 5), but the results were not satisfactory. However, X-ray structures [31] and MD simulations (Figure S3) show that the side chain of Asn364 can adopt different conformations and thus provide extra space for further exploration.

2.2. Hydrogen bond to the side chain of Asp476

The compounds in group 2 feature a heteroaromatic ring substituted with N-methylamide except for compound 12 which has an N-methylthioamide (Table 1). The N-methylamide is involved in hydrogen-bond interactions with the backbone carbonyl group of Ser378 and the side chains of Asn367. Furthermore, their N-methyl groups fit favourably in

![Crystal structure of the complex of YTHDC1 with the m6A nucleoside (PDB 6ZCN).](image)

Fig. 1. Crystal structure of the complex of YTHDC1 with the m6A nucleoside (PDB 6ZCN). (Left) The electrostatic potential at the RNA-recognition surface of YTHDC1 is mainly positive. The electrostatic potential was calculated using the APBS 2.1 plugin of PyMOL [32]. (Right) The detailed view of the binding mode shows the network of hydrogen bonds (green, dashed lines) between m6A (carbon atoms in cyan) and YTHDC1 residues (carbon atoms in grey) or the conserved water molecule (red sphere). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
## Table 1
Chemical structure and binding affinity of 25 ligands of the YTHDC1 m6A-recognition domain.

| Internal Name | 2D structure | Residual signal | IC50 [μM] | PDB code | Resolution [Å] | LE | In-silico Method |
|---------------|--------------|-----------------|-----------|-----------|-----------------|----|-----------------|
| **Group 1**   |              |                 |           |           |                 |    |                 |
| m6A nucleoside| ![m6A nucleoside](image) | 32%             | 504 (56)  | 6ZCN      | 0.23 (0.29)     |    | –               |
| m6A nucleobase| ![m6A nucleobase](image) | 49%             | 918       | –         | 0.38            |    | –               |
| 1             | ![m6A nucleoside](image) | 35%             | 359       | 7P8B      | 0.39            |    | MD              |
| 2             | ![m6A nucleoside](image) | 23%             | 306       | 7P8A      | 0.40            |    | MD              |
| 3             | ![m6A nucleoside](image) | 19%             | 294       | 7P8B      | 0.34            |    | MD              |
| 4             | ![m6A nucleoside](image) | 72%             | –         | –         | –               |    | MD              |
| 5             | ![m6A nucleoside](image) | Nd^d            | –         | –         | –               |    | MD              |
| 6             | ![m6A nucleoside](image) | 3%              | 39 (1.68) | 7P8F      | 0.50 (0.66)     |    | MD              |
| **Group 2**   |              |                 |           |           |                 |    |                 |
| 7             | ![m6A nucleoside](image) | 23%             | 384       | 7PJP      | 0.33            |    | Vina            |
| 8             | ![m6A nucleoside](image) | Nd^d            | –         | 7PJP      | –               |    | Vina            |
| 9             | ![m6A nucleoside](image) | 69%             | –         | 7P87      | –               |    | MD              |
| 10            | ![m6A nucleoside](image) | 90%             | –         | 6YKZ      | –               |    | Vina            |

(continued on next page)
| Internal Name | 2D structure | Residual signal | IC$_{50}$ [μM] | PDB code | Resolution [Å] | LE | In-silico Method |
|---------------|-------------|----------------|----------------|-----------|----------------|----|-----------------|
| 11            | ![2D structure](image1) | 73% | – | 6YKJ | (1.6) | – | Vina |
| 12            | ![2D structure](image2) | 77% | – | 7PJQ | (1.2) | – | SEED |
| 13            | ![2D structure](image3) | 97% | – | 6YNP | (1.1) | – | SEED |
| 14            | ![2D structure](image4) | 64% | – | 7PJ8 | (1.4) | – | Vina |
| 15            | ![2D structure](image5) | Nd$^*$ | – | 7PJ9 | (1.7) | – | Vina |

**Group 3**

| Internal Name | 2D structure | Residual signal | IC$_{50}$ [μM] | PDB code | Resolution [Å] | LE | In-silico Method |
|---------------|-------------|----------------|----------------|-----------|----------------|----|-----------------|
| 16            | ![2D structure](image6) | 84% | – | 6YKI | (1.3) | – | Vina |
| 17            | ![2D structure](image7) | 88% | – | 6YNI | (1.36) | – | Vina |
| 18            | ![2D structure](image8) | 84% | – | 6YL8 | (1.5) | – | Vina |
| 19            | ![2D structure](image9) | 89% | – | 6YKE | (1.52) | – | Vina |
| 20            | ![2D structure](image10) | 74% | – | 6YNJ | (1.5) | – | Vina |
| 21            | ![2D structure](image11) | Nd$^*$ | – | 6YNL | (1.5) | – | Vina |
| 22            | ![2D structure](image12) | 89% | – | 6YL9 | (1.5) | – | Vina |
| 23            | ![2D structure](image13) | 91% | – | 6YNM | (1.5) | – | Vina |

*(continued on next page)*
the aromatic cage (Fig. 2b). The most active compound in this group (compound 7, PDB ID: 7PJ7) shows comparable binding affinity and more favorable ligand efficiency than the m^6A nucleoside (IC_{50} = 384 μM, Figure S4). The two nitrogen atoms of the pyrazole ring of compound 7 form two hydrogen bonds with Asp476 and a structural water molecule [28] (Fig. 2b). Furthermore, its furan ring is involved in a cation-π interaction with the side chain of Arg404, while a similar interaction is not well-defined for the thiopeptide compound 8 (PDB ID: 7PJ7, Figure S2b). The favorable N-methylamide substituent is present in different heteroaromatic rings, but the pyrazole attachment shows the strongest binding affinity in this group. Previously, we reported an N-methylamide substituted indazole with similar potency (compound 26 of Ref. [31]) which shows similar interactions as the ligands in group 2.

### 2.3. Improving binding and ligand efficiency by merging hydrogen-bonding interactions

Compound 7 and m^6A form four hydrogen bonds with the aromatic cage. Three of these hydrogen bonds are in common, i.e., with residues Ser378, Asn367, and the conserved water molecule (Fig. 2). In addition, compound 7 acts as a hydrogen bond donor to the side chain of Asp476. Inspired by the crystal structures, we designed compound 6, which was predicted to form the four hydrogen bonds mentioned above (Fig. 3). We then used ChemAxon [34] to predict the dominant tautomers at pH 7.4. The tautomeric state in which the pyrazole ring can form two hydrogen bonds is the most populated tautomer (with an equilibrium dissociation constant K_d = 0.5 μM) [30]. Importantly, the ligand efficiency improved from 0.29 (m^6A nucleoside) to 0.66 for compound 6, which is even comparable to that of the m^6A nucleoside and compound 7 (Fig. 3). Compared to the m^6A nucleoside, compound 6 shows an approximately 15-fold lower IC_{50} value and 30-fold improvement in the equilibrium dissociation constant measured by ITC (1.7 kcal mol^{-1} vs. 56 kcal mol^{-1}). The binding affinity of compound 6 is even comparable to that of the RNA sequence motif GG(m^6A)CU (K_d = 0.5 μM) [30].

### 2.4. Enlarging the binding pocket by morpholine compounds

The morpholine series shows a displacement of the recognition loop by about 2–5 Å for the Cα atom at the tip of the loop, i.e., residue Gly433 (Fig. 2c and Group 3 in Table 1). The loop opening enlarges the binding site’s volume, thereby permitting more space for further exploration, which is illustrated in the structural overlap of the structures 6ZCN and 6YNI (Figure S7). We then used MarvinSketch to calculate the population of the protomers [34]. In solution, the morpholine is positively charged with a probability ranging from 5% for compound 16 to 90% for compound 27. In the charged state, the morpholine ring forms a cation-π interaction with the aromatic cage, and its methyl or dimethyl substituents are deeply buried in the binding cage. The benzene ring extending from the aromatic cage and the backbone carbonyl of Ala434 (Fig. 2c).

### Table 1 (continued)

| Internal Name | 2D structure | Residual signal | IC_{50} [μM] | PDB code | Resolution [Å] | LE | In-silico Method |
|---------------|--------------|----------------|-------------|-----------|----------------|----|------------------|
| 24            | ![Structure](image) | 88% | – | 7PJ7 | (1.9) | – | SEED |
| **Group 4**   |              |                |             |           |                |    |                  |
| 25            | ![Structure](image) | 85% | – | 6YNK | (1.3) | – | Vina |
| 26            | ![Structure](image) | 56% | – | 6YQ | (1.3) | – | SEED |
| 27            | ![Structure](image) | 74% | – | 7PJ1A | (1.9) | – | SEED |

*Residual HTRF signal in the presence of the small-molecule ligand at a concentration of 1 mM. The lower the residual signal (in percentage) the higher the affinity of the small-molecule ligand. There is no evidence of binding for compounds 4 and 5.*

*IC_{50} values derived from the HTRF competition assay are reported for the compounds whose residual activities at 1 mM compound concentration are smaller than 50%. The values in parentheses for the m^6A nucleoside and compound 6 are the K_d values measured by ITC.*

*The ligand efficiency (LE) values are calculated according to the equation LE = - ΔG_{mol} / n_{HA} = - RT ln(k^+_{HA}) / n_{HA} \approx - RT ln(k^+_{HA}) / n_{HA} , where ΔG is binding free energies from ITC (unit in kcal/mol), n_{HA} number of heavy atoms, IC_{50} from the HTRF assay, R molar gas constant, and T temperature in Kelvin (298 K in this calculation). Most LE values are calculated by substituting IC_{50} for K_d values. For the m^6A nucleoside and compound 6 the LE values were calculated using both IC_{50} value and K_d and the latter is shown in parentheses. Different in-silico methods were used for screening or optimizing the compounds, including MD simulations and automatic docking by AutoDock Vina [37] or SEED [38,39].

*Nd = Not determined because of interference or solubility issues.*




Fig. 2. Representatives of the four groups of ligands (PDB codes 6ZCN, 7PJP, 6YNK, 6YOQ from left to right). (Top row) The binding modes from the crystal structures show the hydrogen bonds (green, dashed lines) between residues in the m6A-recognition pocket (carbon atoms in grey) and the ligands (carbon atoms in cyan). The conserved water molecule is indicated by a red sphere. The water molecule is not shown in panels c and d as it is not present in the complexes with compounds 17 and 25, respectively. These pictures were prepared using the software PyMOL [32]. (Bottom row) Chemical structures. Table 1 contains the chemical structures of all ligands and their binding affinity for the YTH domain of YTHDC1. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
to the morpholine by attaching an ethanesulfonamide group (compound 22). The linked compound roughly maintains the original pose but seems to suffer high ligand strain energy in the sub-pocket (PDB ID: 6YL9). The strain explains, at least in part, why compound 22 does not show an improved binding. Optimizing the linker length could be an option for releasing the high strain energy. Flexible substituents on the N atom of the morpholine ring were designed to reach the Asp476 side chain but did not show any affinity improvement, e.g., compounds 16, 21, and 23. The flexible substituents of these compounds cannot be modeled in the electron density, indicating that less effective interactions exist between them and the solvent-exposed region of the binding site.

2.5. Replacing the structural water molecule by uracil compounds

Our previous study showed that replacing the conserved water molecule (that acts as bridges for hydrogen bonds with the side chains of Trp377 and Asp476) with uracil-containing ligands results in approximately 2 kcal mol\(^{-1}\) penalty in binding free energy [28]. Nonetheless, the uracil derivatives (e.g., compound 25) show promising interactions with YTHDC1. First, the uracil moiety forms five hydrogen bonds with surrounding residues, providing binding specificity (Fig. 2d). Second, substitution is possible for its free carbon atom position, thereby permitting the exploration of the sub-pocket which C11 of compound 25 points to.

If compound 25 is protonated on the tertiary amino group, its positive charge can form a cation-π interaction with the aromatic cage (Figure S8). In addition, the amino group could act as a donor for a hydrogen bond with the carbonyl oxygen of Ser378. However, we cannot examine this possibility solely based on the crystal structure. We thus resorted to in-silico methods. The population of the protonated state was predicted as 38% in aqueous solution compared to 62% for the neutral form, and the strain energy of the ligand in the bound pose is small for both forms (by OpenEye's Freeform) [35,36]. Multiple MD simulations show substantially higher structural stability for the neutral state than the charged species by comparing their pose RMSD values, approximately 1 Å vs. 5 Å, respectively (Figure S8). This simulation result is consistent with the charge state of the natural ligand, as m^6^A does not bear any formal charge. Therefore, further optimization on this series shall aim to stabilize the neutral protomer rather than focusing on a positively charged (tertiary) amino group.

Compound 26 shows moderate binding affinity (56% residual signal at 1 mM) with YTHDC1 and offers a possibility for further optimization. This compound captures most of the interactions which m^6^A does with YTHDC1 (Fig. 2e). For example, the 2-pyridone core forms hydrogen bonds with Ser378 and the structural water molecule. The aromatic cage recognizes the 6-methyl group. By analyzing its crystal structure (PDB ID: 6YOQ), we find that replacing C5 of the 2-pyridone ring with a nitrogen atom may introduce an extra hydrogen bond to Asn367. Moreover, the solvent-exposed sulfonamide group does not show specific interactions with the reader domain. Therefore, it could be used as a linker for growing fragments to the solvent-exposed pocket.

3. Conclusion

In summary, we have identified 25 small-molecule ligands of the YTH domain of YTHDC1 by in silico methods, i.e., molecular docking and molecular dynamics simulations. Their crystal structures in the complex with YTHDC1 were solved at high resolution (1.2–1.9 Å). The SAR of the ligands and their interactions with YTHDC1 were used to suggest directions for hit optimization. The merging of two weak ligands, namely the m^6^A nucleobase (IC\(_{50}\) = 900 μM) and compound 7 (IC\(_{50}\) = 400 μM), has resulted in the ligand-efficient compound 6 (IC\(_{50}\) = 39 μM) which has \(K_d\) = 1.7 μM as measured by ITC and a ligand efficiency of 0.66 kcal mol\(^{-1}\) n\(_{HA}\). Furthermore, compound 6 does not show significant binding to the off-target YTHDF3.

4. Materials and methods

4.1. Docking-based virtual screening (VS)

Two docking engines were used to screen small-molecule libraries in this study, namely, Autodock Vina 1.1.2 [37] and SEED [38,39]. In this virtual screening campaign, our focus was to efficiently search for
potential hit compounds for further experimental validation. Protein flexibility was not taken into account despite its potential usefulness [40].

The structure of YTHDC1 in complex with the oligoribonucleotide GG(m^6A)ACU (PDB ID: 4RI3) [27] was used for the docking campaign. The structure was prepared for docking using the same procedure as for the MD simulations (see below). After the minimization, the oligoribonucleotide, all the water molecules, and ions were removed from the system. In the VS stage by docking, we decided to remove all solvent molecules to allow for a large variety of putative binding modes. Interestingly, the docking programs suggested also ligands that replace the conserved water molecule (Fig. 2d). The continuum electrostatics treatment of the solvent (see below) approximates the water molecules, including the conserved ones, by a high-dielectric region which results in a favorable contribution for polar groups of the ligand. On the other hand, the presence of water molecules would not allow for the placements of ligand atoms because of van der Waals repulsion.

For the screening with AutoDock Vina, the binding site was defined by a 30 × 30 × 30 Å cubic grid box centered on the geometric center of the m^6A nucleoside coordinates. We set the parameter “exhaustiveness” to 10 to explore the configurational space of ligands and left the rest of the parameters as the default. Only the best-scoring poses were saved for further analysis. About 0.4 million molecules from the ZINC’s fragment sub-chemical database [41] were screened. After docking with Vina, we only considered the top 1000 molecules ranked by the predicted ligand efficiency, i.e., docking score divided by the number of heavy atoms. The final potential hit list was pruned by discarding compounds with less than three hydrogen bonds with YTHDC1.

The docking program SEED with its polar docking mode was used to screen small (Q<20) and mainly rigid molecules (about 0.1 million from ZINC). The fragments were placed by SEED with at least one hydrogen bond to the backbone polar groups or side chain of residue Ser378 which was considered as the center of the binding site. The residues within 15 Å of Ser378 were taken into account for the calculations of electrostatic and van der Waals energy terms. The specifications of ligand atoms for the binding site definition are optional in SEED, and in this case, it was not necessary because Ser378 is deeply buried in the m^6A binding site. SEED uses an efficient procedure for the numerical calculation of the Born radii [42,43] which are used for the evaluation of the generalized Born electrostatic energy terms. The docking with SEED approximately required 1–5 s for each molecule, while the docking with AutoDock Vina required about 20 s.

4.2. Pose validation by MD simulations

All systems were prepared based on the m^6A-YTHDC1 complex X-ray structure (PDB ID: 6ZCN) [28]. Chain A of the structure was used as the template to construct other complex systems. The protein (residue 345 to 506), ligand (i.e., m^6A nucleoside), and crystal water molecules were mainly kept while the other buffer components were removed. Protonation states of all ionizable residues were determined based on their pKa values in water and interaction environments with the protein. Missing atomic coordinates were built using the IC tables of CHARMM36 force parameters [44,45]. Each system was solvated with a 70 Å rhombic decahedron (RHDO) water box to ensure a 12 Å buffer water layer between the system and 150 mM with Na^+ to mimic physiological conditions, we set the ionic concentration of the decahedron (RHDO) water box to ensure a 12 Å buffer water layer between the system and 150 mM with Na^+ to mimic physiological conditions, we set the ionic concentration of the decahedron (RHDO) water box to ensure a 12 Å buffer water layer between the system and 150 mM with Na^+. The system was prepared by a high-dielectric region which results in a favorable contribution for polar groups of the ligand. On the other hand, the presence of water molecules would not allow for the placements of ligand atoms because of van der Waals repulsion.

For the screening with AutoDock Vina, the binding site was defined by a 30 × 30 × 30 Å cubic grid box centered on the geometric center of the m^6A nucleoside coordinates. We set the parameter “exhaustiveness” to 10 to explore the configurational space of ligands and left the rest of the parameters as the default. Only the best-scoring poses were saved for further analysis. About 0.4 million molecules from the ZINC’s fragment sub-chemical database [41] were screened. After docking with Vina, we only considered the top 1000 molecules ranked by the predicted ligand efficiency, i.e., docking score divided by the number of heavy atoms. The final potential hit list was pruned by discarding compounds with less than three hydrogen bonds with YTHDC1.

The docking program SEED with its polar docking mode was used to screen small (Q<20) and mainly rigid molecules (about 0.1 million from ZINC). The fragments were placed by SEED with at least one hydrogen bond to the backbone polar groups or side chain of residue Ser378 which was considered as the center of the binding site. The residues within 15 Å of Ser378 were taken into account for the calculations of electrostatic and van der Waals energy terms. The specifications of ligand atoms for the binding site definition are optional in SEED, and in this case, it was not necessary because Ser378 is deeply buried in the m^6A binding site. SEED uses an efficient procedure for the numerical calculation of the Born radii [42,43] which are used for the evaluation of the generalized Born electrostatic energy terms. The docking with SEED approximately required 1–5 s for each molecule, while the docking with AutoDock Vina required about 20 s.

4.3. Protein expression, purification, and X-ray crystallography

Recombinant YTH domain of YTHDC1 protein comprising residues 345–509 (YTHDC1 345–509) was expressed in E. coli and purified as described in detail previously [30]. In summary, the N-terminally hexahistidine-tagged YTH domain was purified using immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography. Crystal structures of the YTH domain in complex with the compounds were obtained by soaking compounds into the crystals of YTHDC1 as reported before [31].

4.4. Binding assays

The residual signals at 1 mM compound concentration and the IgG values were obtained through an HTRF-based assay as previously described [31,33]. The assay evaluates the binding interaction of the YTHDC1 YTH domain (amino acids 345–509) and a methylated RNA (sequence: 5’-Biotin-AAGACCCG(m^6A)CUAGCCU-3’). The YTH domain of YTHDC1 is expressed as a GST-fusion protein that is recognized by an anti-GST antibody labelled with Eu^3+, acting as the Förster resonance energy transfer (FRET) donor. The biotinylated RNA is bound by Streptavidin conjugated to XL665, the FRET acceptor. The binding of the methylated RNA to the YTH binding site leads to the formation of a four-member complex constituted by the GST-tagged YTH domain of YTHDC1, anti-GST Eu^3+-labelled antibody, biotinylated RNA, and Streptavidin conjugated to XL665. This complex formation subsequently leads to the proximity between the FRET donor and acceptor, resulting in a signal emission. Suppose the tested compound can compete with the methylated RNA for the occupation of the YTH active site, the emitted signal decreases. The assay mix includes 25 nM YTHDC1 YTH domain-GST fusion protein, 15 nM biotinylated RNA (Dharmacon), 0.8 nM anti-GST Eu^3+-conjugated antibody (Cisbio, 61GSTKLB), 1.875 nM XL665-conjugated streptavidin (Cisbio, 610SAXLB), and the compound of interest. The compound concentration in the final mix strictly depends on the assay’s aims. It is fixed at 1 mM to investigate the residual signal or as an array of 1 mM 2-fold dilutions to determine the IgG50 values. The assay’s components are diluted in a buffer composed of 50 mM HEPES pH 7.5, 150 mM NaCl, 100 mM kF, and 0.1% BSA. The reagents mix is incubated for 3 h at 24 °C, transferred into a white, low volume 384-well
4.5. Chemistry

All reagents were used as received unless otherwise noted. The solvents were dried over activated molecular sieves of appropriate size. All reactions were monitored by TLC or NMR. Chromatography was carried out over silica gel by hand or automated flash chromatography (Büchi C-850 flash prep).

1H and 13C NMR spectra were recorded on AV2 400 MHz Bruker spectrometer. Chemical shifts are given in ppm. The spectra are calibrated to the residual 1H and 13C signals of the solvents. Multiplicities are abbreviated as follows: singlet (s), triplet-triplet (tt), multiplet (m), and broad (br). Melting points were determined on a Büchi M – 560 melting point instrument. High-resolution electrospray ionization mass spectrometry was performed on an Agilent 1290 Infinity LC system coupled to an Agilent 6540 quadrupole time-of-flight mass spectrometer. The jet stream electrospray source was operated in positive mode with the following parameter settings: nebulizer pressure 35 psi, nozzle voltage 0 V, sheath gas flow 11 L/min, sheath gas temperature 375 °C, drying gas flow 8 L/min, drying gas temperature 250 °C, capillary voltage 3000 V and fragmentor voltage of 175 V. Accurate mass spectra were acquired in profile mode over an m/z range of 100–1000 by 1 spectrum per second. The Q-TOF instrument was operated in high-resolution mode with 1700 m/z instrument mass range at a resolving power of 33 000 (measured at m/z 322). The purity was acquired by HPLC on an Agilent LC device using a NUCLEOSHELL RP18 column (50 × 4.6 mm, 2.7 μm). The eluent at flow rate of 1.5 mL/min consisted of MeCN and 0.01 M (NH4)2SO4 pH = 6.6 as solvents. All compounds showed >95% purity. MarvinSketch [34] was used for drawing, displaying, and characterizing chemical structures and protonation states as well as for generating conformers.

4.5.1. N9,9-Dimethyl-9H-purin-6-amine (compound 2)

To a stirred suspension of 6-chloro-9-amine-9H-purine (100 mg, 0.59 mmol) in isopropanol (2 mL), methylamine (3 equiv., 1.78 mmol, 8 M in EtOH) was added. The reaction mixture was stirred at 80 °C for 2 h in the microwave and concentrated under reduced pressure. The crude residue was purified by column chromatography (EtOAc/heptane = 9:1 to EtOAc/MeOH = 100:5 to 100:20) to afford the desired product as a yellow solid (83 mg, 88% yield).1H NMR (400 MHz, CDCl3): δ = 8.54 (s, 1 H), 7.72 (s, 1 H), 5.84 (br s, 1 H), 3.83 (s, 3 H), 3.23 (br s, 3 H); 13C NMR (101 MHz, CDCl3): δ = 155.4, 153.1, 149.1, 139.9, 119.6, 29.6, 27.4; HRMS (ESI): m/z: calcld for [C8H8N4O2]+: 179.1045 found: 179.1040.

4.5.2. 6-Chloro-N4-cyclopropylpyrimidine-4,5-diamine (intermediate I of compound 3)

To a stirred solution of 6-chloropyrimidine-5-amine (1 g, 6.1 mmol) in isopropanol (12 mL), cyclopropylamine (2 equiv., 12.2 mmol, 849 μL) and Et3N (2 equiv., 12.2 mmol, 1.7 mL) were added. The reaction mixture was stirred at 100 °C for 6 h in the microwave and concentrated under reduced pressure. The obtained residue was poured into water and the aqueous layer was extracted three times with EtOAc. The combined organic layers were washed with brine, dried over MgSO4, filtered and concentrated under reduced pressure to afford the desired product as a brown solid (1.11 g, 98% yield).1H NMR (400 MHz, CDCl3): δ = 8.18 (s, 1 H), 5.14 (br s, 1 H), 3.38 (br s, 2 H), 2.92–2.82 (m, 1 H), 0.93–0.87 (m, 2 H), 0.62–0.56 (m, 2 H); 13C NMR (101 MHz, CDCl3): δ = 156.0, 149.7, 143.0, 122.0, 24.2, 7.2 (s, 2 C).

4.5.3. 6-Chloro-9-cyclopropyl-9H-purine (intermediate I2 of compound 3)

To a stirred solution of aniline intermediate II (1.1 g, 5.6 mmol) in triethylthiophosphorane (14 mL), Et3O (12 μl) was added. The reaction mixture was stirred at 25 °C for 1 h and concentrated under reduced pressure. The crude residue was purified by column chromatography (EtOAc/heptane = 3:7 to 1:3) to afford the desired product as a white solid (78 mg, 84% yield).1H NMR (400 MHz, CDCl3): δ = 8.79 (s, 1 H), 8.13 (s, 1 H), 3.57–3.50 (m, 1 H), 1.32–1.16 (m, 4 H); 13C NMR (101 MHz, CDCl3): δ = 152.8, 151.9, 150.8, 145.7, 131.7, 25.6, 5.9 (s, 2 C); HRMS (ESI): m/z: calcld for [C8H11ClN4]+: 193.0950 found: 193.51.

4.5.4. 9-Cyclopropyl-N-methyl-9H-purin-6-amine (compound 3)

To a stirred suspension of chloropurine intermediate I2 (100 mg, 0.51 mmol) in isopropanol (1.7 mL), methylamine (3 equiv., 1.54 mmol, 193 μl, 8 M in EtOH) was added. The reaction mixture was stirred at 80 °C for 2.5 h in the microwave and concentrated under reduced pressure. The crude residue was purified by column chromatography (EtOAc/heptane = 1:1 to 7:3 to 10:0 to 100:10) to afford the desired product as a white solid (98 mg, quantitative yield). HPLC Purity = 100%; Mp: 159–162 °C; 1H NMR (400 MHz, CDCl3): δ = 8.47 (s, 1 H), 7.74 (s, 1 H), 5.70 (br s, 1 H), 3.43 (tt, J = 7.2, 3.6 Hz, 1 H), 3.22 (br s, 3 H), 1.24–1.16 (m, 2 H), 1.15–1.08 (m, 2 H); 13C NMR (101 MHz, CDCl3): δ = 155.4, 153.2, 150.0, 139.9, 119.9, 27.4, 24.9, 5.7 (s, 2 C); HRMS (ESI) m/z: calcld for [C8H11N3Cl]+: 190.1093 found: 190.1087.

4.5.5. N2,N6-Dimethyl-9H-purine-2,6-diamine (compound 4)

To a stirred suspension of 2,6-dichloro-9H-purine (100 mg, 0.49 mmol) in isopropanol (1.63 mL), methylamine (6 equiv., 2.96 mmol, 369 μl, 8 M in EtOH) was added. The reaction mixture was stirred at 150 °C for 4 h in the microwave and concentrated under reduced pressure. The crude residue was purified by column chromatography (DCM/MeOH = 102:2 to 100:4 to 100:10) to afford the desired product as a white solid (27 mg, 28% yield). HPLC Purity = 100%; Mp: decomposed; 1H NMR (400 MHz, MeOD-d4): δ = 7.65 (s, 1 H), 3.06 (s, 3 H), 2.92 (s, 3 H); 13C NMR (101 MHz, DMSO-d6): δ = 160.6, 155.5, 152.0 (br s, 1 C), 135.3, 113.6 (br s, 1 C), 28.8, 27.3 (br s, 1 C); HRMS (ESI) m/z: calcld for [C11H11N2]: 179.0854 found: 179.0840.

4.5.6. N2,N9,9-Trimethyl-9H-purine-2,6-diamine (compound 5)

To a stirred suspension of 2,6-dichloro-9-methyl-9H-purine (200 mg, 0.99 mmol) in isopropanol (3 mL), methylamine (5 equiv., 4.93 mmol, 616 μl, 8 M in EtOH) was added. The reaction mixture was stirred at 150 °C for 8 h in the microwave and concentrated under reduced pressure. The crude residue was purified by column chromatography (DCM/MeOH = 100:0 to 100:1 to 100:2 to 100:3) and the impure product was triturated in diethyl ether with a few drops of acetone, filtered and washed with diethyl ether to afford the desired product as a white solid (115 mg, 61% yield). HPLC Purity = 100%; Mp: 228–229 °C; 1H NMR (400 MHz, MeOD-d4): δ = 7.62 (s, 1 H), 3.67 (s, 3 H), 3.05 (s, 3 H), 2.95 (s, 3 H); 13C NMR (101 MHz, DMSO-d6): δ = 161.0, 155.1, 151.2, 137.4, 113.3, 28.8, 28.3, 26.8; HRMS (ESI) m/z: calcld for [C14H16N2]: 213.1202 found: 213.1196.

Accession number

The atomic coordinates and structure factors for the 25 crystal
structures have been deposited in the Protein Data Bank (PDB) with the accession codes 7P88, 7P8A, 7P8B, 7P8F, 7PJP, 7PJ7, 7PJ8. 6YKZ, 6YKJ, 7PQJ, 6YNP, 7PJ8, 6YKI, 6YNI, 6LYE, 6YNE, 6YNL, 6YL9, 6YNM, 7PB6, 6YNK, 6YOQ, and 7PJA.

Declaration of competing interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Beat Blattmann at the Protein Crystallization Center (University of Zurich) for assistance with crystallization screening. We thank Dr. Lars Wiedmer for help with some of the biochemical assays. The use of beamlines and user support at Swiss Light Source are gratefully acknowledged. We thank the Swiss National Supercomputing Center (CSCS) in Lugano and the Swedish National Institute for Computing (SNIC) at the High-Performance Computing Center North (HPC2N), partially funded by the Swedish Research Council through grant agreement no. 2018-05973, for providing the computational resources. This work was supported by the International Postdoc Grant funded by the Swedish Research Council (Grant VR 2019-00608 to Y.L.) and an Excellence grant of the Swedish National Science Foundation (SWEN10030B 189363 to A.C.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmcrr.2022.100057.

References

[1] C. He, Grand challenge commentary: RNA epigenetics? Nat. Chem. Biol. 6 (12) (2010) 863-865.
[2] K.D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C.E. Mason, S.R. Jaffrey, Rethinking m6A readers, writers, and erasers, Annu. Rev. Genet. 53 (2019) 1-321.
[3] J. Liu, J. Li, L. He, J. Wen, H. Ma, B. Song, J. Diao, L. Wang, J. Wang, F. Wu, L. Tan, Y.G. Shi, Y. Shi, H. Shen, METTL3 regulates histone H3 in mouse embryonic stem cells, Nature 591 (7849) (2021) 321-326.
[4] J. Liu, M. Gao, J. He, K. Wu, S. Lin, L. Yin, C. Chen, H. Liu, J. Shi, X. Wang, L. Chang, Y. Wang, X.L. Zhao, Z. Xie, Z. Yang, G.Z. Luo, G. Wu, P. Ji, W. Xiong, J. Ren, The RNA m6A reader YTHDC1 silences retrotransposons and guards ES cell identity, Nature 591 (7849) (2021) 322-326.
[5] H. Shima, M. Matsumoto, Y. Ishijima, M. Ebina, A. Muto, S. Kamagai, K. Ochiai, T. Suzuki, K. Ishikawa, C. Hata, N6-Methyladenosine methylation is regulated by selective N6-methyladenosine demethylase and m6A degradation involving METTL16 and YTHDC1, Cell Rep. 21 (12) (2017) 3354-3363.
[6] Y. Cheng, X. Wei, B.F. Pickering, K.L. Chou, A.M. Savino, X. Yang, H. Luo, D.T. Nguyen, S. Nguyen, M. Safra, S. Geula, C. Soyris, A. Nachshon, J. Tai, T. Pan, C. He, Dynamic RNA modification and alternative ITS expression regulation, Cell 169 (7) (2017) 1187-1197.
[7] R. Winkler, E. Gillis, L. Lasman, M. Safra, S. Geula, C. Soyris, A. Nachshon, J. Tai, T. Pan, C. He, Dynamic m6A methylation directs translational control of heat shock response, Nature 526 (7574) (2015) 173-182.
[8] Y. Li, R.K. Bedi, L. Wiedmer, S. Adhikari, U. Dahal, Y.S. Chen, Y.J. Hao, B.F. Sun, H.Y. Sun, A. Li, X.L. Ping, W.Y. Lai, X. Wang, H.L. Ma, C.M. Huang, Y. Yang, N. Guang, G.B. Jiang, H.L. Wang, Q. Zhou, X.J. Wang, Y.L. Zhao, Y.G. Yang, Nuclear m6A reader YTHDC1 regulates mRNA splicing, Mol. Cell 61 (4) (2016) 507-519.
[9] I.A. Roundtree, G.Z. Luo, Z.J. Zhang, X. Wang, T. Zhou, Y.Q. Cui, J.H. Sha, Y.N. Huang, Guerrero, P. F, E. He, B. Shen, C. He, YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs, Elife 6 (2017) e1311.
[10] D.P. Patil, C.K. Chen, B.F. Pickering, A. Chow, C. Jackson, M. Guttman, S.R. Jaffrey, m6A RNA methylation promotes XIST-mediated transcriptional repression, Nature 537 (7620) (2016) 369-373.
[11] R.K. Bedi, D. Huang, L. Wiedmer, Y. Li, A. Dolbois, J.A. Wojdyla, M.E. Sharpe, M.G. Kharas, Methylated RNA regulates alternative splicing and phase separated nuclear body that suppresses myeloid leukemic differentiation, Cancer Cell 39 (7) (2021) 958-972, e2.
[12] R. Shi, S. Ying, Y. Li, Z. Zhu, X. Wang, H. Jin, Linking the YTH domain to cancer: the importance of YTH family proteins in epigenetics, Cell Death Dis. 12 (4) (2021) 364.
[13] 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. Comm. 10 (11) (2019) 925-932.
[14] Y. Li, R.K. Bedi, L. Wiedmer, X. Sun, D. Huang, A. Caflisch, Atomatic and thermodynamic analysis of N6-methyladenosine (m6A) recognition by the reader domain of YTHDC1, J. Chem. Theor. Comput. 17 (2) (2021) 1240-1249.
[15] M. Krepl, F.F. Damberger, C. von Schroetter, D. Theler, P. Pokorna, F.H. Allain, A. Caflisch, HMGB1 is a methylated RNA binding protein that regulates stress response in breast cancer cells, J. Mol. Cell. Biol. 33 (2017) 319-342.
[16] Y.Z. Li, R.K. Bedi, L. Wiedmer, D.Z. Huang, P. Sledz, A. Caflisch, Flexible binding of m6A reader protein YTHDC1 to its preferred RNA motif, J. Chem. Theor. Comput. 15 (12) (2019) 7004-7014.
[17] R.K. Bedi, D. Huang, L. Wiedmer, Y. Li, L. Dobos, J.A. Wojdyla, M.E. Sharpe, A. Caflisch, P. Sledz, Selectively disrupting m6A-dependent protein-RNA interactions with fragments, ACS Chem. Biol. 15 (3) (2020) 618-625.
[18] The PyMOL Molecular Graphics System, Version 2.2, Schrodinger, LLC.
[19] L. Wiedmer, S.A. Eberle, R.K. Bedi, P. Sledz, A. Caflisch, A reader-based assay for m6A writers and erasers, Anal. Chem. 91 (4) (2019) 3078-3084.
[20] MarvinSketch 21.20.0, ChemAxon.http://www.chemaxon.com.
[21] S. Wlodek, A.G. Skillman, A. Nicholls, Ligand entropy in gas-phase, upon solvation and in explicit solvent simulations: potentials of mean force and conformational entropy, J. Chem. Theor. Comput. 51 (5) (2015) 1127-1135.
[22] M. Krepl, F.F. Damberger, C. von Schroetter, D. Theler, P. Pokorna, F.H. Allain, A. Caflisch, Comparison of a GB solvation model with GB/SA, used in molecular simulations: potentials of mean force and conformational entropy, J. Chem. Theor. Comput. 15 (12) (2019) 7004-7014.
[23] K. Ochiai, T. Suzuki, K. Igarashi, S-adenosylmethionine synthesis is regulated by the YTHDC1 domain of YTHDC1, J. Chem. Theor. Comput. 17 (2) (2021) 1240-1249.
[24] D.T. Nguyen, S. Mo, E. Barin, A. Velleca, T.M. Rohwetter, D.J. Patel, S.R. Jaffrey, Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain, Nature 591 (7849) (2021) 322-326.
[43] M. Scarsi, J. Apostolakis, A. Caflisch, Continuum electrostatic energies of macromolecules in aqueous solutions, J. Phys. Chem. 101 (43) (1997) 8098–8106.

[44] Y. Xu, R. Vanommeslaeghe, A. Aleksandrov, A.D. Mackerell, L. Nilsson, Additive CHARMM force field for naturally occurring modified ribonucleotides, J. Comput. Chem. 37 (10) (2016) 896–912.

[45] J. Huang, A.D. Mackerell, CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, J. Comput. Chem. 34 (25) (2013) 2135–2145.

[46] Maestro 11.5, S., LLC, New York, NY, 2018.

[47] G.J. Martyna, M.L. Klein, M. Tuckerman, Nose-hoover chains - the canonical ensemble via continuous dynamics, J. Chem. Phys. 97 (4) (1992) 2635–2643.

[48] S.E. Feller, Y.H. Zhang, R.W. Pastor, B.R. Brooks, Constant-pressure molecular-dynamics simulation - the Langevin piston method, J. Chem. Phys. 103 (11) (1995) 4613–4621.

[49] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell Jr., CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, J. Comput. Chem. 31 (4) (2010) 671–690.

[50] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (19) (1995) 8577–8592.

[51] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, J.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (10) (2009) 1545–1614.