Supporting information

Fragment ligands of the m^{6}A-RNA reader YTHDF2

Francesco Nai[a], Raed Nachawati[a], František Zálešák[a], Xiang Wang[a], Yaozong Li[a,b,*], and Amedeo Caflisch[a,*]

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

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

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

Table of Contents

Materials and methods

Supporting schemes, figures and tables

Materials and methods

1. Protein purification

The version of the YTHDF2 YTH domain used for the crystallographic experiments (His_6-tagged, residues 408-552) was overexpressed in Rosetta (DE3) cells, overnight at 18 °C. The expression was induced with IPTG at a final concentration of 0.2 mM. The cells were lysed by sonication in a buffer composed of 20 mM Hepes pH 7.0 and 500 mM NaCl. The lysate was clarified by centrifugation for one hour at 48,000 g, 4 °C. All the subsequent steps were performed at 4 °C. The clarified lysate was loaded into an HisTrap high performance 5 mL column (Cytiva). The loading step was followed by a washing step using a buffer composed of 20 mM Hepes pH 7.0, 500 mM NaCl and 20 mM Imidazole. After the washing step, the protein was eluted using a buffer composed of 20 mM Hepes pH 7.0, 500 mM NaCl and 250 mM Imidazole. The obtained protein solution was dialyzed overnight in a buffer composed of 20 mM Hepes pH 7.0 and 150 mM NaCl. The protein solution was subsequently loaded in a HiTrap Heparin HP 5 mL column (Cytiva) and eluted using a linear gradient of NaCl up to a final concentration of 1 M. The protein was further purified with a gel filtration step using a HiLoad 16/50 Superdex 75 pg column (Cytiva) and a buffer composed by 20 mM Hepes pH 7.0, 150 mM NaCl and 5 mM DTT.

The version of the YTHDF2 YTH domain used for the HTRF-based binding assays (GST-tagged, residues 383-579) was overexpressed in Rosetta (DE3) cells, overnight at 18 °C. The expression was induced with IPTG at a final concentration of 0.2 mM. The cells were lysed by sonication in a buffer composed of 50 mM Na2HPO4 pH 7.2, 150 mM NaCl, 1 mM DTT, 1 mM PMSF and 1 mM EDTA. The clarified lysate was loaded on Glutathione Sepharose 4B resin (GE Healthcare). The loading step was followed by a washing step using a buffer composed of 50 mM Na2HPO4 pH 7.2, 150 mM NaCl, 1 mM DTT and 1 mM EDTA. After the washing step the protein was eluted using a buffer composed of 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT and 10 mM reduced Glutathione. The protein was further purified with a gel filtration step using a HiLoad 16/600 Superdex 200 pg column (Cytiva) and a buffer composed of 10 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM DTT.

The same protocol was applied for the purification of YTHDC1 (GST-tagged, residues 345-509) YTHDF1 (GST-tagged, residues 361-559) and YTHDF3 (GST-tagged, residues 391-585).

2. Protein crystallography and soaking

We launched a systematic screening of crystallization conditions to optimize apo crystals of the m^6A-reader domain of YTHDF2 (residues 408-552) for soaking. About 700 crystallization conditions were tested by varying buffer, precipitant, pH, and salt conditions. It emerged that YTHDF2 crystals can grow under several conditions. However, the crystal packing may be too tight in some of them, which prevents

ligand soaking. We eventually identified a small subset of conditions (ammonium sulfate 2M, sodium citrate tribasic pH 5.4) for determining the structure of YTHDF2 in complex with m^6A. The crystals were subsequently improved by using a 6x4 matrix of conditions defined by an Ammonium sulfate concentration range of 1.925 M-2.05 M and by a pH range of 4.9-5.2 in presence of Sodium citrate tribasic at a fixed concentration of 0.1 M. The protein was mixed with the mother liquor solutions in a 1:1 ratio to a final concentration of 7.5 mg/mL. The trays were set up using the hanging drop vapor diffusion technique at 4 °C. Crystals were observed in 5 days and reached full maturation in 10 days.

Soaking was performed by inserting 1 µL of saturated fragment solution directly into the crystallization drops and incubating overnight at 4 °C. The saturated fragment solution was prepared by dissolving the fragments in a solution composed of 0.1 M Sodium citrate tribasic pH 5.0 and 2 M Ammonium sulfate and separating the non-solubilized fragment excess through centrifugation. After overnight incubation, the crystals were passed through a cryoprotectant solution constituted by a 20% glycerol mother liquor solution and flash-frozen in liquid nitrogen.

The X-ray diffraction experiment was performed on the X06DA beamline of Paul Scherrer institute's Swiss Light Source. The collected data were reduced using XDS[1] and scaled using XSCALE, the structures were solved using PHASER[2] and the structure of YTH-YTHDF2 in complex with m^6A (Ref. 3 PDB ID: 4RDN) as a search model. The search model was prepared by eliminating water molecules and m^6A from the structure. Model rebuilding and refinement were performed using COOT[4] and Phenix.refine[5] respectively.

3. HTRF-based binding assay

The GST-tagged version of the YTHDF2 YTH domain was used to determine the IC50 values and the residual signal at 1 mM fragments concentration, similarly as previously reported.[6] The HTRF assay used in this study is based on the binding of the YTH domain of YTHDF2 to a RRACH-containing methylated oligoRNA (where R stands for G/A, and H for ACU, Ref. 7) with sequence 5'-AAGAACCGG(m^6A)CUAAGCU-3. A fluorescent signal is emitted when the protein binds to the oligoRNA and the signal is decreased when a binding competitor is introduced. We first tested the compounds in the HTRF assay at a single dose of 1 mM concentration except for the poorly soluble compounds. The dose-response measurement for IC50 determination was carried out for the compounds that showed a signal reduction of at least 50% with respect to DMSO control. The single dose measurements and the individual data points of the dose-response measurements are the average of two or more biological replicates, each of which is the average of two technical replicates.

The assay mix is constituted of four components at the following final concentrations: 25 nM YTHDF2 YTH domain-GST fusion protein, 15 nM biotinylated RNA (Dharmacon), 0.8 nM anti-GST Eu3+-conjugated antibody (Cisbio), 1.875 nM XL665-conjugated streptavidin (Cisbio). The assay buffer in

which the reagents are diluted is composed of 50 mM HEPES pH 7.5, 150 mM NaCl, 100 mM KF and 0.1% BSA. The fragment of interest is also added to the mix and tested either at the single concentration of 1 mM or as a set of 2-fold dilutions where the starting concentration is normally 1 mM. The assay mix is assembled in 96 well-plate, incubated for 3 hours at room temperature and transferred to a white, low volume 384-well plate (Corning) before measurement. Emission at 620 and 665 nm is measured after excitation at 620 nm and applying an excitation/emission lag time of 60 µs. The measurement is carried on with a Tecan Infinite M1000 plate reader. The dose-response data are analyzed using GraphPad Prism 9.

The same protocol was applied for the selectivity testing on YTHDC1, YTHDF1 and YTHDF3.

The assay used in this work can be considered a simplified and direct version of the METTL3-14 reader-based assay. The latter was characterized as High-throughput screening compatible[6] and is therefore reasonable to assume the same for the direct version. Moreover, the METTL3-14 reader-based assay is characterized by a reported assay window of 9.7 (Ref. 6)while the direct version features a normally better assay window that can reach values higher than 30 in optimal conditions. The assay is characterized by low variability and high reproducibility as shown in figure S1 which depicts the single biological replicates used to produce the averaged dose-response curves of figure 5.

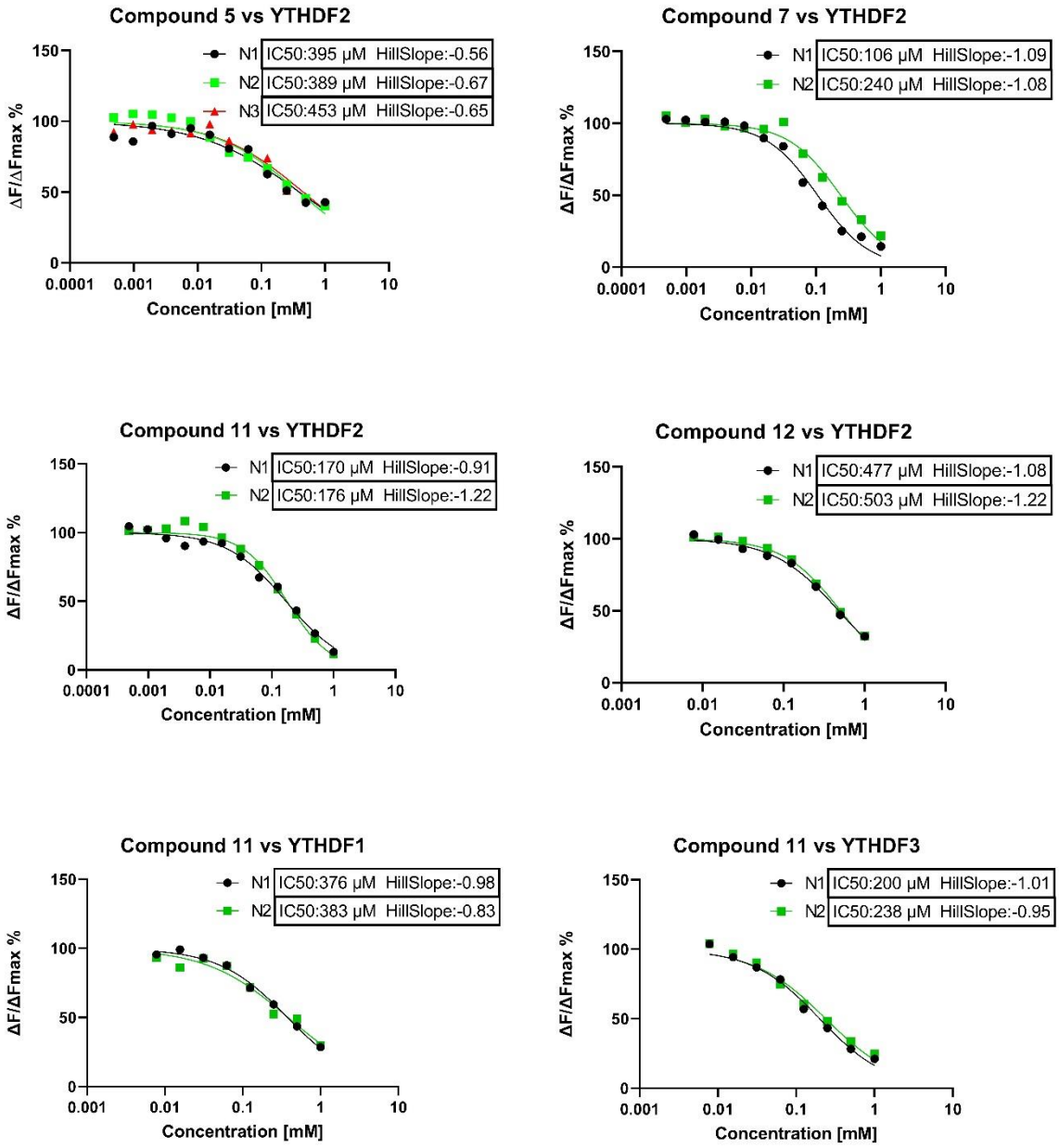


Figure S1. HTRF dose-response curves. The biological replicates are displayed separately. Each data point is the average of two technical replicates.

4. Validation of designed compounds by molecular dynamics (MD) simulations

The simulation systems were prepared based on the X-ray structure of the complex with ligand **9** which was used as the template to design 15 compounds. The YTHDF2 reader domain (residue 408 to 549) and crystal water molecules were mostly kept while the other buffer components were removed. Protonation states and orientations of residue side chains were determined based on their pKa values in water and interaction environments with the protein. The most populated tautomers and protomers of designed compounds were determined with MarvinSketch 21.20.0, ChemAxon,

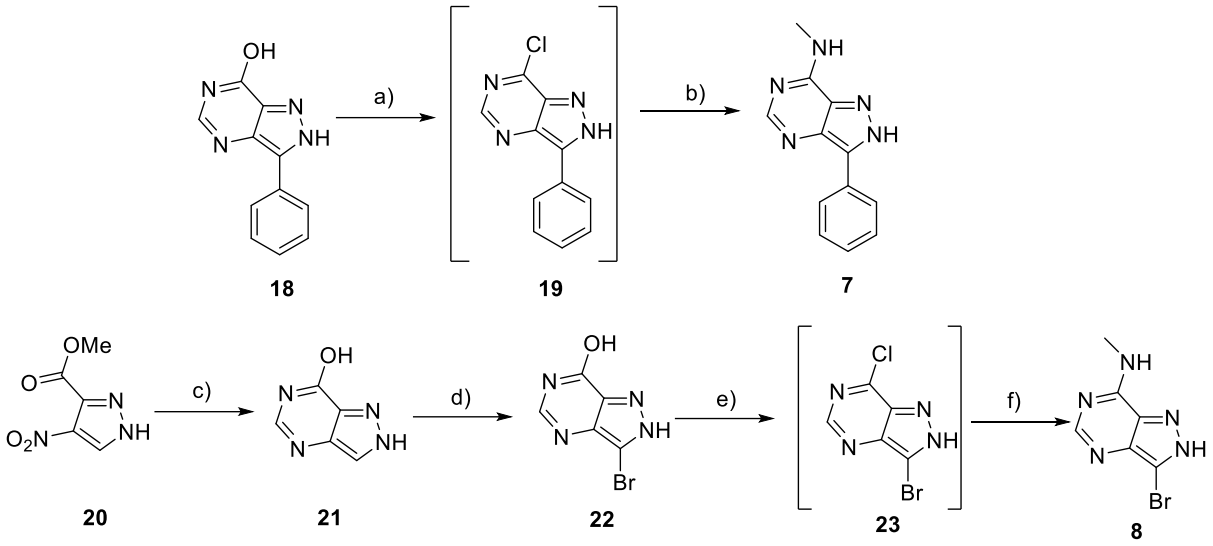
(http://www.chemaxon.com). Missing protein coordinates were built using the IC tables of CHARMM36 force field parameters.[8] Each system was solvated with a 68 Å rhombic dodecahedron water box to ensure a 12 Å buffer layer between the protein atoms and the boundary of the water box. The protocol of the MD simulations is the same as we published recently.[9] For each of the 15 compounds, five independent simulations were started with different initial assignments of the velocities, and each replicate lasted for 20 ns. The binding pose of a bound ligand during MD simulations was compared to its originally designed pose to calculate its non-hydrogen atom root-mean-square-deviation (RMSD). Nine of 15 compounds showed stable poses (average RMSD value below 3 Å) and were purchased through Chemspace (https://chem-space.com/). The RMSD time series of two representative compounds are displayed in Figure S3.

5. *Chemistry*

The tested compounds comprise both commercially available and in-house synthesized compounds **(2-4, 7, 8)**. Synthesis of compounds **2-4** was previously reported by our research group.[9] Here we report synthesis of compounds **7** and **8.** Compound **7** was synthesized from 3-phenyl-1,6-dihydro-7H-pyrazolo[4,3-d]pyrimidin-7-one **18** that was prepared based on previously reported method.[10] Compound **8** was prepared with modified procedure from commercially available methyl 4-nitro-1H-pyrazole-5-carboxylate **20**. The remaining compounds were acquired through Chemspace (www.chem-space.com) which guaranteed purity ≥95%.

All reagents were purchased from commercial suppliers and used as received. Reactions run at elevated temperature were carried out in the oil bath. All reactions were monitored by thin-layer chromatography (Aluminium plates coated with silica gel 60 F_{254}). Flash column chromatography was carried out over silica gel (0.040-0.063 mm). ^{1}H and ^{13}C {^{1}H} NMR spectra were recorded on AV2 400 MHz and AV600 Bruker spectrometers (400 MHz, 101 MHz and 600 MHz, 150 MHz, respectively) in DMSO or CD_3OD. Chemical shifts are given in ppm and their calibration was performed to the residual ^{1}H and ^{13}C signals of the deuterated solvents. Multiplicities are abbreviated as follows: singlet (s), doublet (d) multiplet (m), and broad signal (bs). The purity was acquired by Liquid chromatography high resolution electrospray ionization mass spectrometry (LC-HR- ESI-MS): Acquity UPLC (Waters, Milford, USA) connected to an Acquity eλ diode array detector and a Synapt G2 HR-ESI-QTOF-MS (Waters, Milford, USA); injection of 1 µL sample (c = ca. 10-100 µg mL−1 in the indicated solvent); Acquity BEH C18 HPLC column (1.7 µm particle size, 2 × 50 mm, Waters) kept at 30 °C;* elution at a flow rate of 400 µL min−1 with A: H2O + 1% HCO2H and B: CH3CN + 0.1% HCO2H, linear gradient from 5–98% B within 5 min, then isocratic for 1 min;* UV spectra recorded from 200–600 nm at 1.2 nm resolution and 20 points s−1; ESI: positive ionization mode, capillary voltage 3.0 kV, sampling cone 40V, extraction cone 4V, N2 cone gas 4 L h−1, N2 desolvation gas 800 L min−1, source temperature 120°C; mass

analyzer in resolution mode: mass range 100–2'000 m/z with a scan rate of 1 Hz; mass calibration to <2 ppm within 50–2'500 m/z with a 5mM aq. soln. of HCO2Na, lockmasses: m/z 195.0882 (caffein, 0.7 ng mL−1) and 556.2771(Leucine-enkephalin, 2 ng mL−1).



Scheme S1. a) $POCl_3$, *N*,*N*-dimethylaniline, 100 °C, 3 h; b) 33 % $MeNH_2$ in EtOH, EtOH, reflux, 2 h, 6 % (after two steps); c) (i) Pd/C 10%, $HCOONH_4$, MeOH, reflux, 55 °C, 1 h; (ii) formamidine acetate, 2-ethoxyethanol, reflux, 14 h, 64 % (after two steps); d) NBS, DMF, rt, 17 h, 77 %; e) $POCl_3$, *N*,*N*-dimethylaniline, 100 °C, 2 h; f) 33% $MeNH_2$ in EtOH, EtOH, reflux, 2 h, 31 % (after two steps).

N-methyl-3-phenyl-1H-pyrazolo[4,3-d]pyrimidin-7-amine (7)

3-phenyl-1H-pyrazolo[4,3-d]pyrimidin-7-ol **18** (0.063g, 0.29 mmol) was dissolved in $POCl_3$ (0.6 mL, 6.4 mmol) followed by addition of *N*,*N*-dimethylaniline (0.12 mL, 0.94 mmol). The reaction was heated to 100 °C for 3 hours. The reaction was cooled to rt and volatiles were removed *in vacuo*. The residue was extracted into Et_2O (3x5 mL), combined organic layers were dried over $MgSO_4$ and evaporated. The crude product was purified using flash column chromatography (SiO_2; EtOAc:Pentane – 1.5 :1). Impure 7-chloro-3-phenyl-1H-pyrazolo[4,3-d]pyrimidine **19** was dissolved in EtOH (0.5 mL) and 33% $MeNH_2$ in EtOH was added (0.3 mL). The reaction mixture was refluxed for 2 hours. After the reaction completion, the volatiles were removed under reduced pressure. The crude product was purified using flash column chromatography (SiO_2; EtOAc) and isolated as white solid (4 mg, 6 % after two steps). ^{1}H NMR (600 MHz, CD_3OD- *d*$_4$) δ 8.31 (s, 1H), 8.17 (d, J = 7.7 Hz, 2H), 7.51 – 7.47 (m, 2H), 7.42 – 7.37 (m, 1H), 3.16 (s, 3H). ^{13}C {^{1}H} NMR (101 MHz, CD_3OD- *d*$_4$) δ 152.0, 128.4, 128.1, 126.5, 26.4 (heteroaromatic carbons missing). LRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{12}H_{12}N_5$; 226.109 found, 226.109.

1H-pyrazolo[4,3-d]pyrimidin-7-ol (21)

To a suspension of 10% Pd/C (0.08 g) in MeOH (5 mL) was added methyl 4-nitro-1H-pyrazole-5-carboxylate **20** (500 mg, 2.92 mmol). The reaction mixture was heated to 55 °C and $HCOONH_4$ (1.285 g, 20.44 mmol) was carefully added portion wise (dihydrogen evolution!!!). The reaction mixture was stirred at the same temperature for additional 1 hour. After consumption of starting material (TLC), the reaction mixture was cooled to rt and Pd/C was filtered off. After the filtrate evaporation, the resulting amine was dissolved in 2-ethoxyethanol (6.75 mL). Formamidine acetate (0.91 g, 8.76 mmol) was added and the reaction mixture was refluxed for 14 h. After cooling the reaction mixture down, DCM (10 mL) and heptane (10 mL) was added. The formed precipitate was filtered off and suspended in MeOH (10 mL). The suspension was refluxed for 1 hour and the beige solid was collected by filtration (0.254 g, 64 % after two steps). ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.11 (s, 1H), 7.85 (s, 1H); ^{13}C {^{1}H} NMR (101 MHz, DMSO-d_6) δ 154.2, 143.0, 138.9, 131.1, 129.3; LRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_5H_5N_4O$; 137.046 found, 137.046.

3-bromo-1H-pyrazolo[4,3-d]pyrimidin-7-ol (22)

1H-pyrazolo[4,3-d]pyrimidin-7-ol **21** (0.25 g, 1.8 mmol) was suspended in DMF (2 mL) followed by addition of NBS (0.357 g, 2 mmol). The reaction was stirred 17 h at rt and DCM (2 mL) was added after the reaction completion. The solid was filtered off and the filtration cake was washed with DCM (2x2.5 mL). The product **22** in the form of grey solid was dried on air at room temperature (0.295 g, 77 %). %). ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.07 (s, 1H); ^{13}C {^{1}H} NMR (101 MHz, DMSO-d_6) δ 152.7, 144.0, 143.01, 137.54, 128.2; LRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_4H_4BrN_4$; 214.956 found, 214.956.

3-bromo-*N*-methyl-1H-pyrazolo[4,3-d]pyrimidin-7-amine (8)

Compound **22** (0.210 g, 0.98 mmol) was dissolved in $POCl_3$ (2.05 mL, 21.9 mmol) followed by addition of *N,N*-dimethylaniline (0.43 mL, 3.3 mmol). The reaction slurry was refluxed for 2 hours and the volatiles were removed *in vacuo*. The residue was poured over ice-water mixture and extracted into EtOAc (3x5 mL). Combined organic phases were dried over $MgSO_4$, decolorized by charcoal and concentrated under reduced pressure. The residue was purified using flash column chromatography (SiO_2; DCM:MeOH 2:0.3 to 2:0.5). The impure intermediate **23** was directly dissolved in EtOH (1.5 mL) followed by addition of 33% $MeNH_2$ in EtOH (0.9 mL). The reaction mixture was refluxed for 2 hours and after the reaction completion, the solvent was removed *in vacuo*. The crude product was purified using flash column chromatography (SiO_2; DCM:MeOH - 2:0.15 to 2:0.2). Compound **8** was isolated in the form of white solid (0.07 g, 31 % after two steps). ^{1}H NMR (400 MHz, DMSO-d_6) δ 13.11 (bs, 1H), 8.30 (s, 1H), 7.63 (bs, 1H), 3.04 (d, 4.6 Hz, 3H); ^{13}C {^{1}H} NMR (101 MHz, DMSO-d_6) δ 153.0, 150.4, 139.3, 123.1, 121.5, 27.5; LRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_6H_7BrN_5$; 227.988 found, 227.988.

6. *Safety statement*

No unexpected or unusually high safety hazards were encountered.

Methods Bibliography:

(1) Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (2), 125–132. https://doi.org/10.1107/S0907444909047337.

(2) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.

(3) Li, F.; Zhao, D.; Wu, J.; Shi, Y. Structure of the YTH Domain of Human YTHDF2 in Complex with an m(6)A Mononucleotide Reveals an Aromatic Cage for m(6)A Recognition. *Cell Res.* **2014**, *24* (12), 1490–1492. https://doi.org/10.1038/cr.2014.153.

(4) Emsley, P.; Cowtan, K. *Coot* : Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60* (12), 2126–2132. https://doi.org/10.1107/S0907444904019158.

(5) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with *Phenix.Refine*. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (4), 352–367. https://doi.org/10.1107/S0907444912001308.

(6) Wiedmer, L.; Eberle, S. A.; Bedi, R. K.; Śledź, P.; Caflisch, A. A Reader-Based Assay for M6A Writers and Erasers. *Anal. Chem.* **2019**, *91* (4), 3078–3084. https://doi.org/10.1021/acs.analchem.8b05500.

(7) Du, H.; Zhao, Y.; He, J.; Zhang, Y.; Xi, H.; Liu, M.; Ma, J.; Wu, L. YTHDF2 Destabilizes M6A-Containing RNA through Direct Recruitment of the CCR4–NOT Deadenylase Complex. *Nat. Commun.* **2016**, *7* (1), 12626. https://doi.org/10.1038/ncomms12626.

(8) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(9) Li, Y.; Bedi, R. K.; Nai, F.; von Roten, V.; Dolbois, A.; Zálešák, F.; Nachawati, R.; Huang, D.; Caflisch, A. Structure-Based Design of Ligands of the M6A-RNA Reader YTHDC1. *Eur. J. Med. Chem. Rep.* **2022**, 100057. https://doi.org/10.1016/j.ejmcr.2022.100057.

(10) Stephanos, I. VALO HEALTH INC. US2016/229872, 2016, A1.

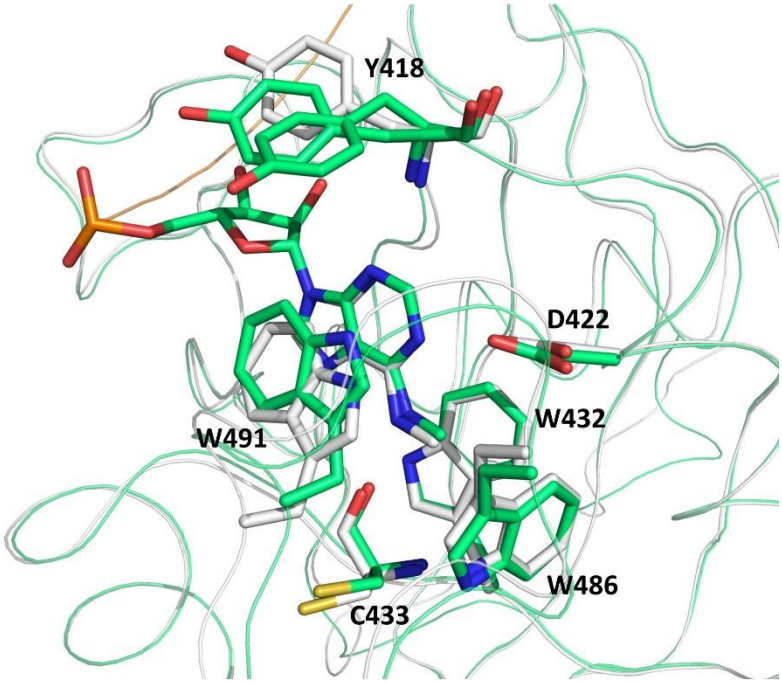


Figure S2. Structural overlap of YTHDF2 in complex with m^6adenine (compound **1**) (white, PDB code: 7YWB) and m^6A containing oligoRNA (limegreen, 7Z26). Oxygen, nitrogen, sulfur and phosphorus atoms are shown in red, blue, yellow and orange, respectively.

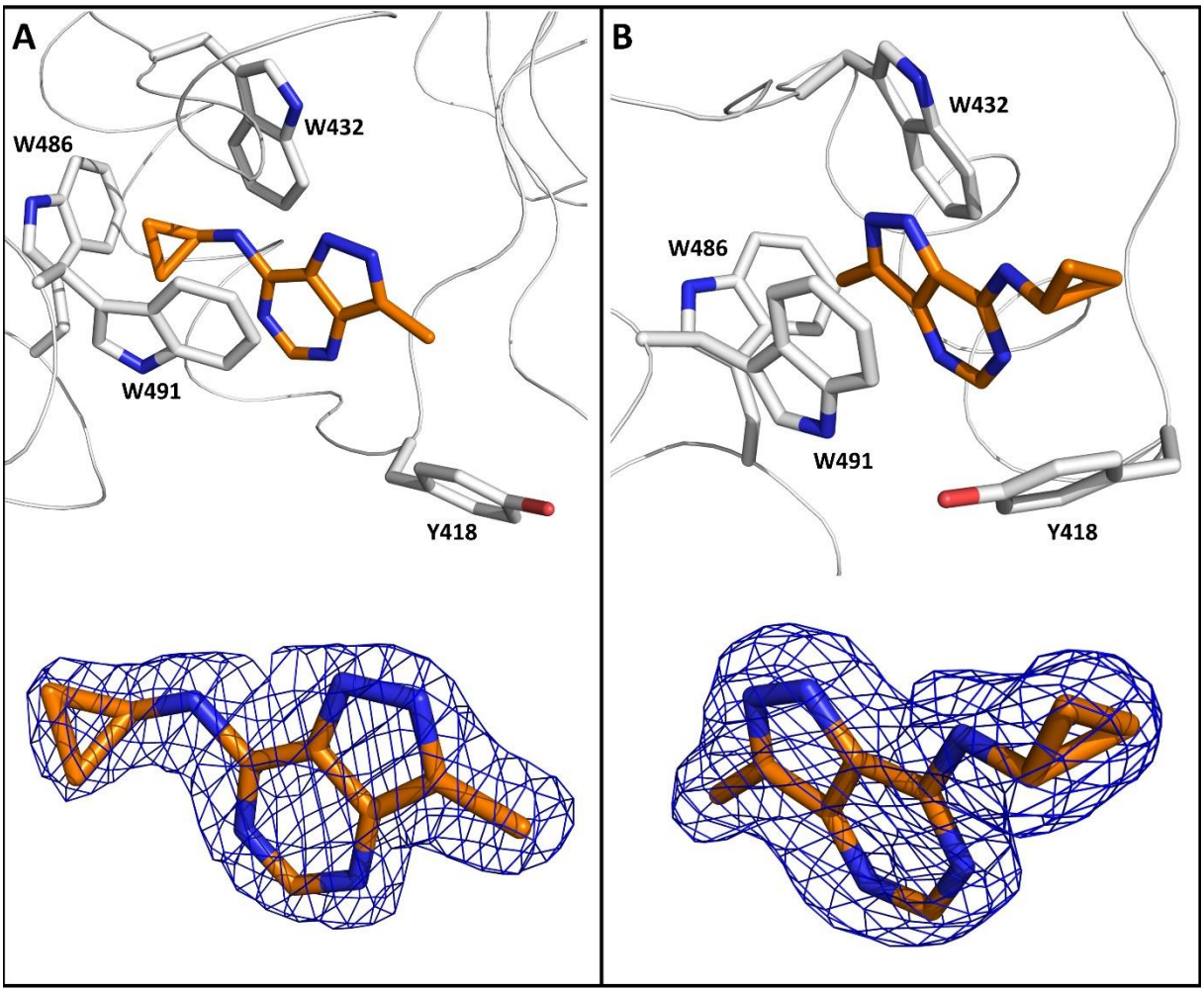


Figure S3. Orientations of compound **6** in the YTHDF2 binding pocket (PDB code: 7Z8W). Compound **6** in YTHDF2 chain B (panel A) and chain A (panel B). The carbon atoms of the ligands are in orange and those of the protein in white. Ligand 2Fo-Fc electron density is displayed at a contour level of 1 σ.

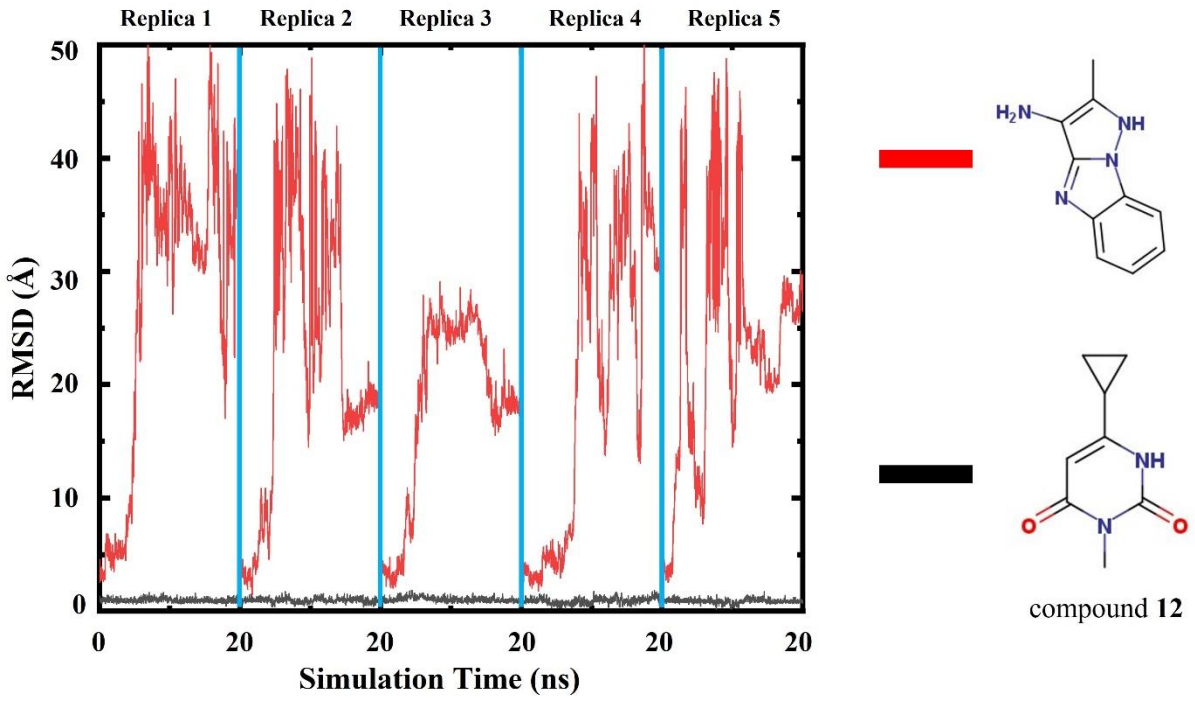


Figure S4. Stability of the predicted pose of potential binders of YTHDF2 during MD simulations. The plot shows the time series of the root mean square deviation (RMSD) of the pose with respect to the manually generated binding mode. The five independent runs for each of the two compounds are separated by vertical lines (cyan). Compound **12** shows a stable pose in all five MD runs (black), while the tricyclic compound dissociates from the binding pocket within the first 5-10 ns (red).

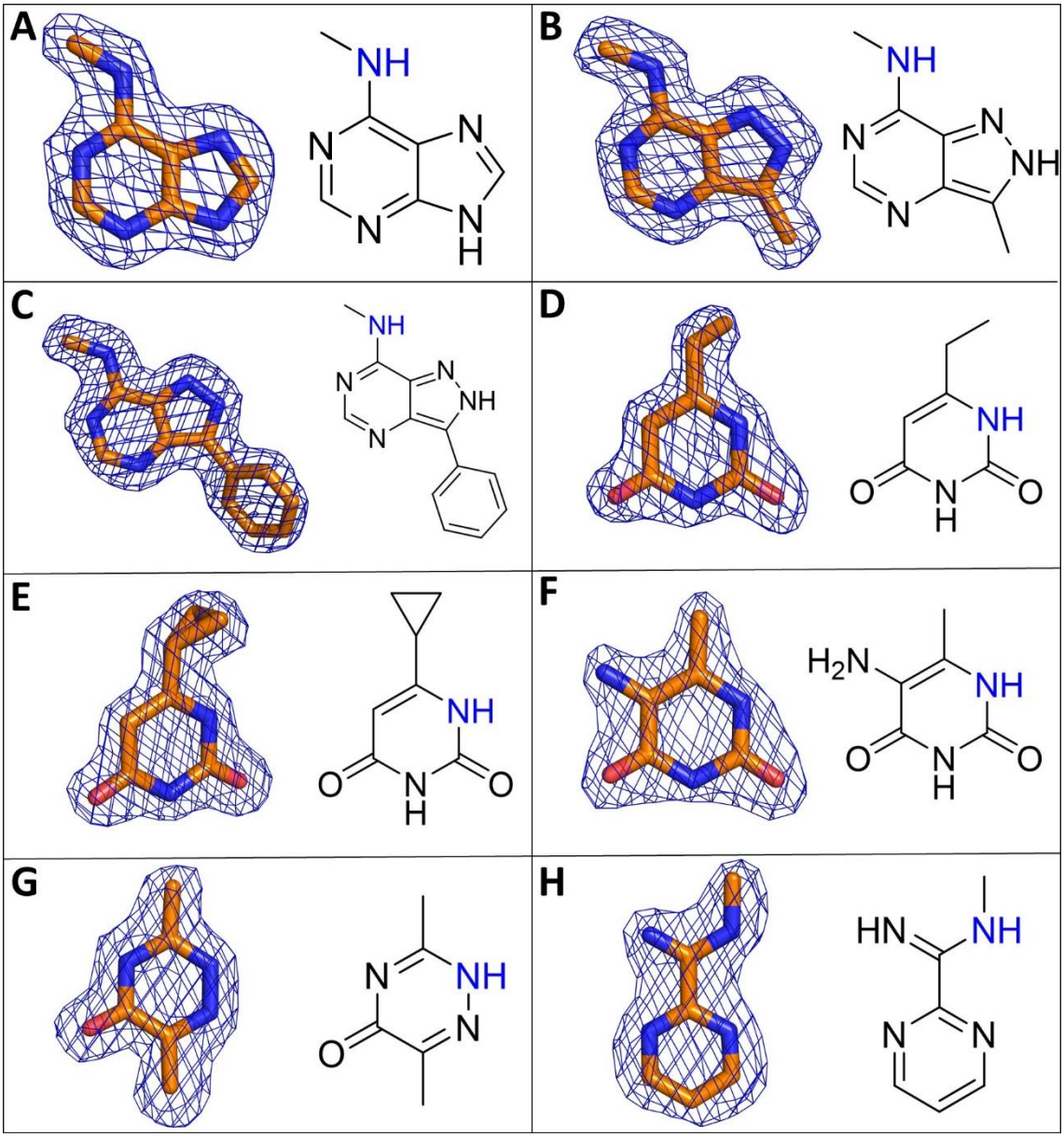


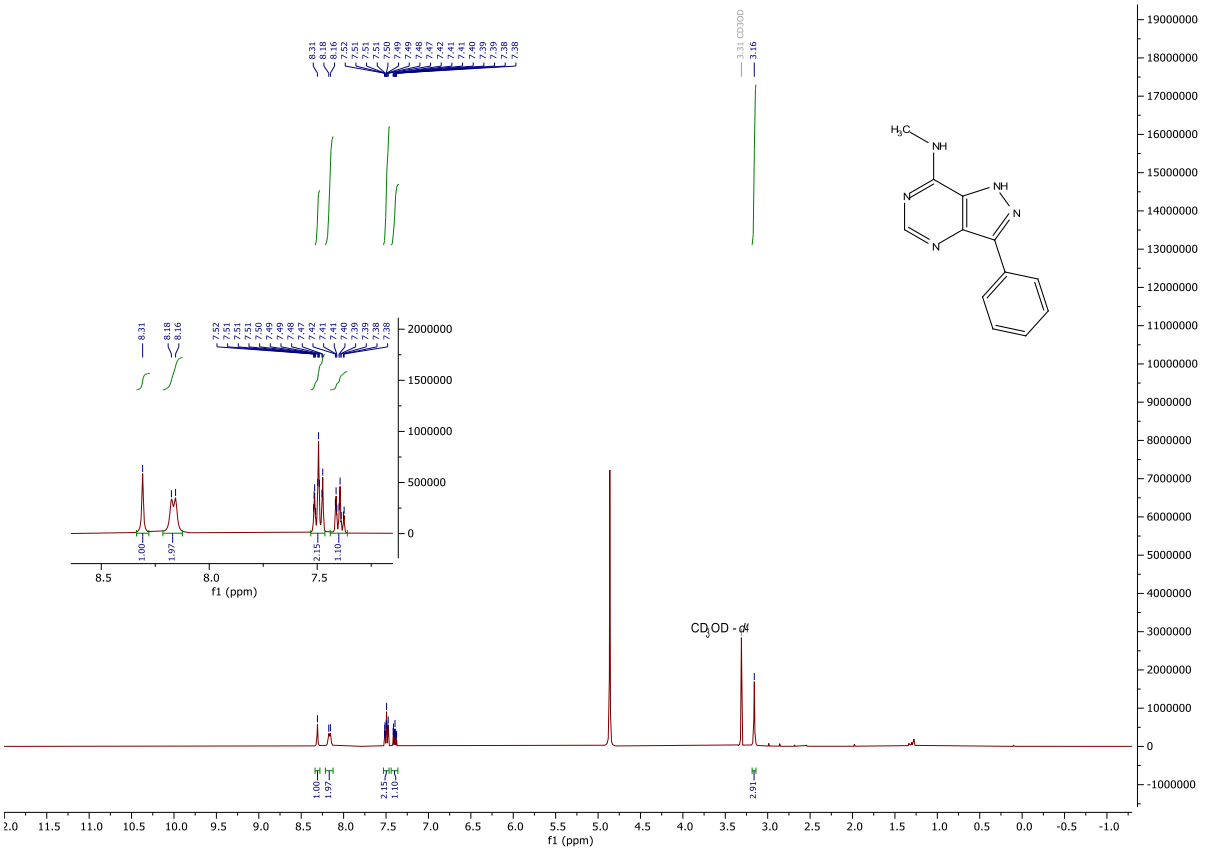
Figure S5. 2Fo-Fc electron density maps (contoured at 1 σ) and 2D structures of eight of the 17 ligands of the YTHDF2 reader domain. (A-H) Compounds **1** (PDB code: 7YWB), **5** (7Z5M), **7** (7YXE), **9** (7YX6). **11** (7R5W), **13** (7R5F), **16** (7R5L), and **17** (7ZG4), respectively.

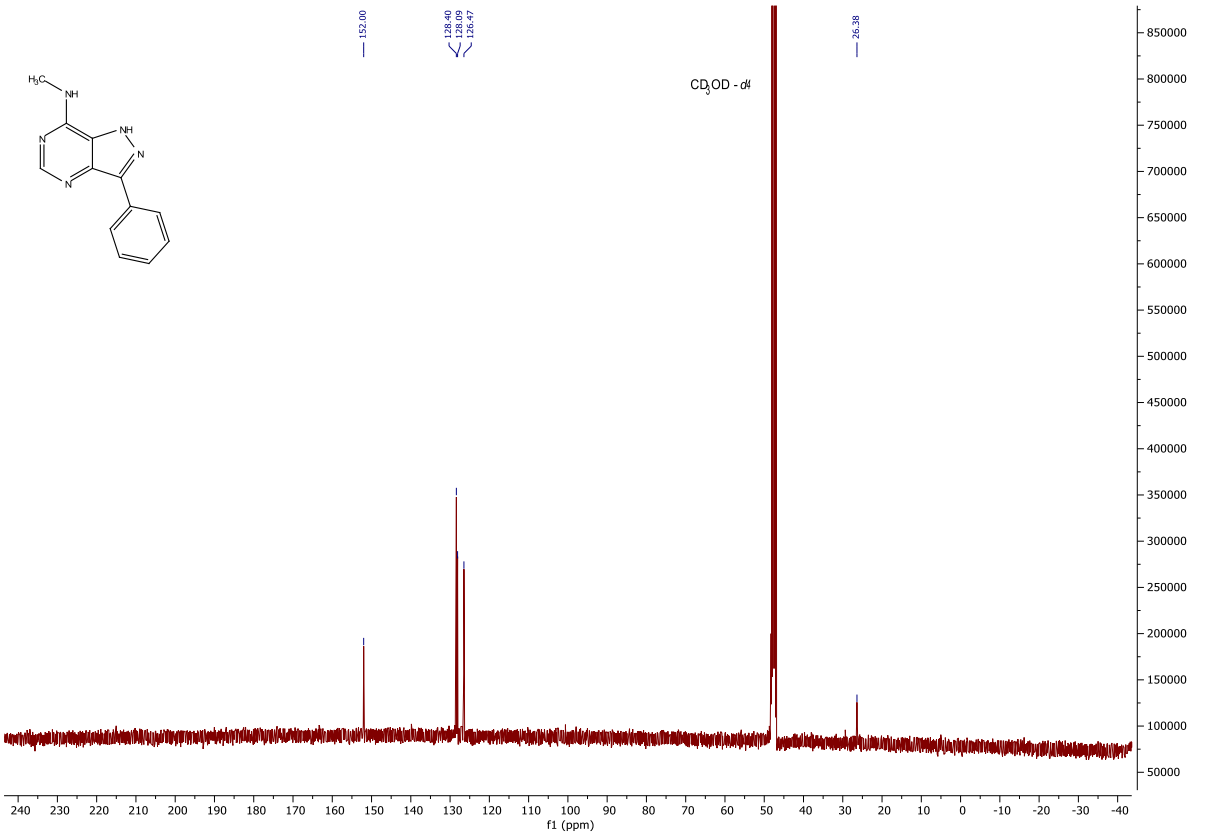
	Residual signal at 1 mM compound concentration [%]		
	YTHDF1	**YTHDF2**	**YTHDF3**
5	67	42	65
7	49	18	59
11	29	12	23
12	60	32	43

Table S1. Selectivity measurements for the four fragments for which the IC50 for YTHDF2 was measured. The reported values are the average of two or more biological replicates, each of which is the average of two technical replicates.

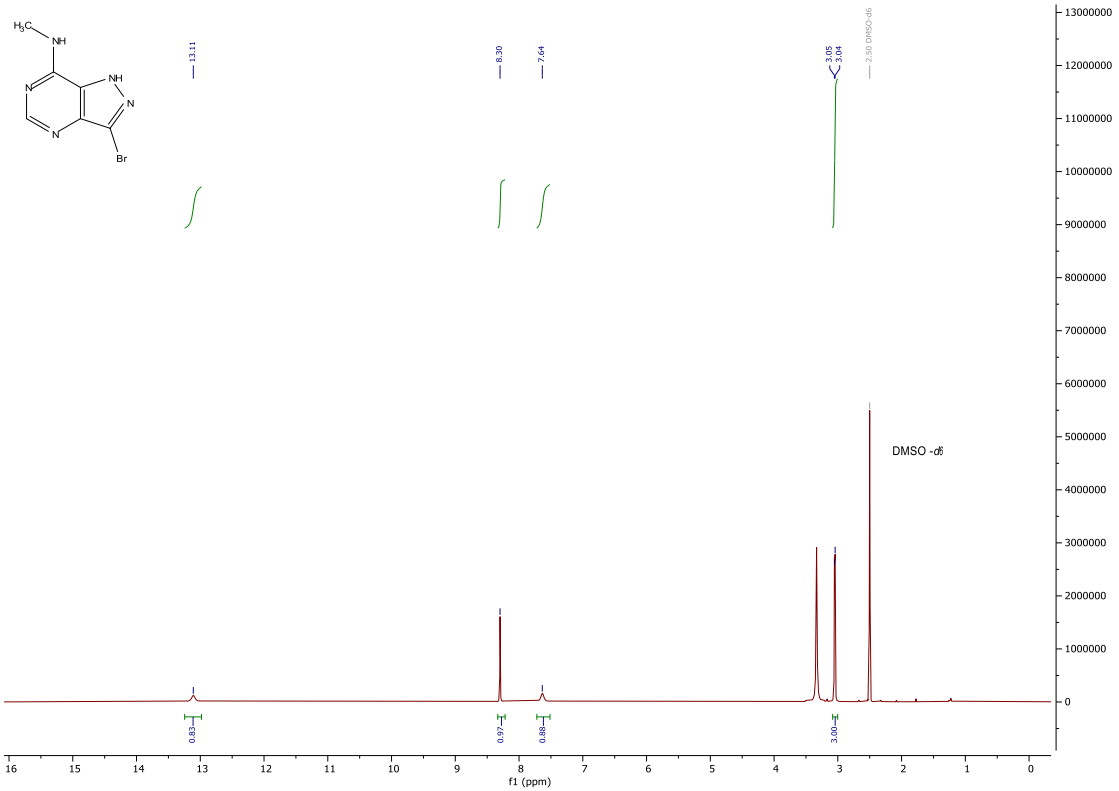
^{1}H and ^{13}C spectra of synthesized final compounds

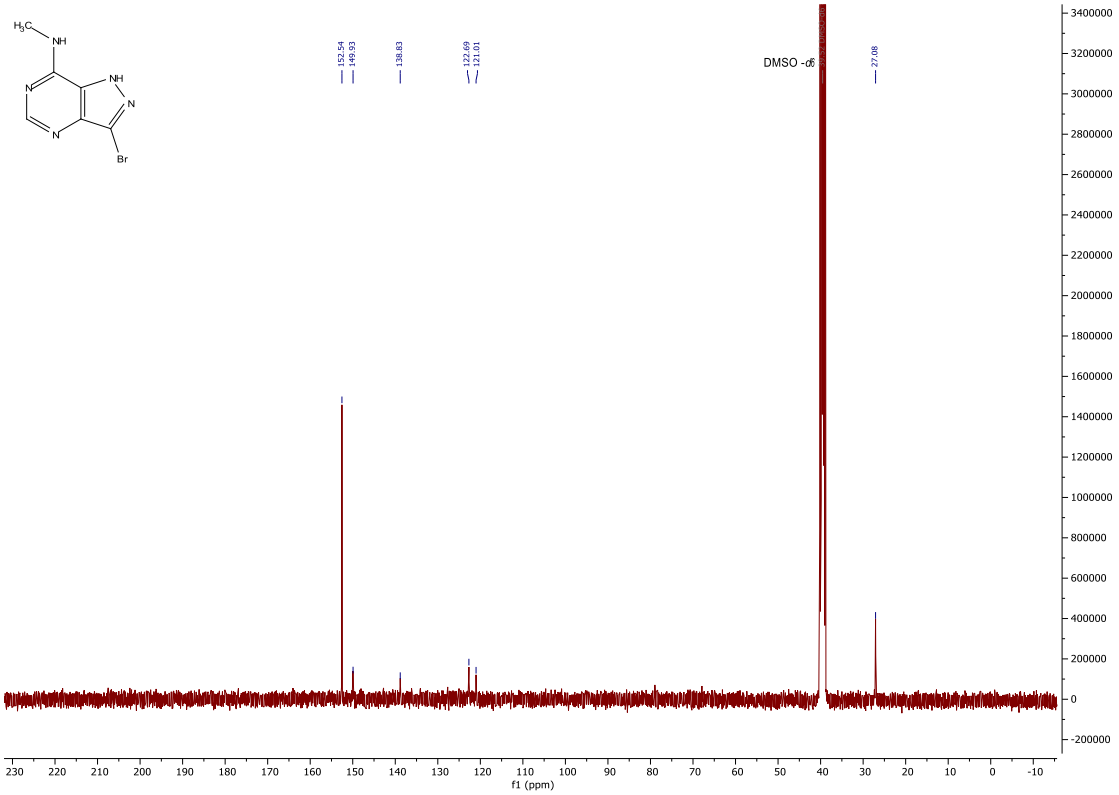
N-methyl-3-phenyl-1H-pyrazolo[4,3-d]pyrimidin-7-amine (7)





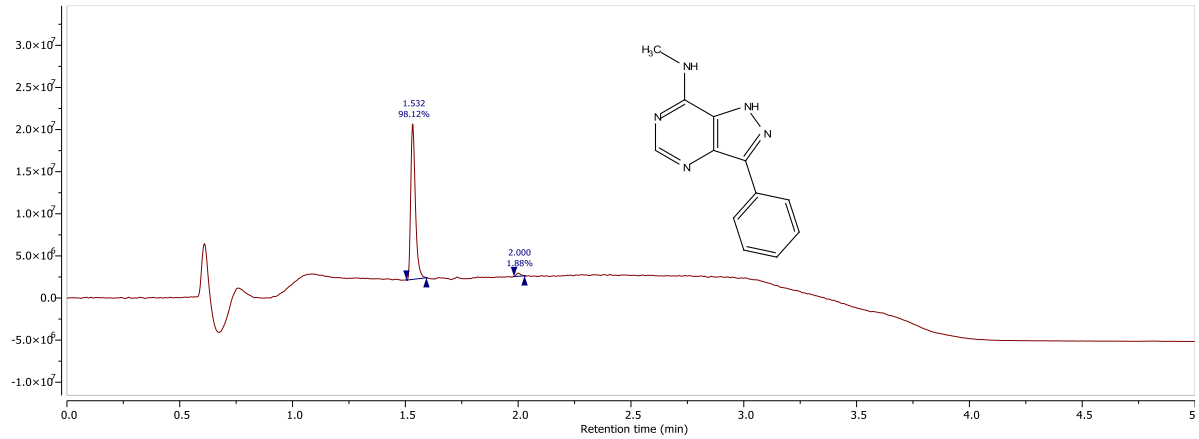
3-bromo-N-methyl-1H-pyrazolo[4,3-d]pyrimidin-7-amine (8):



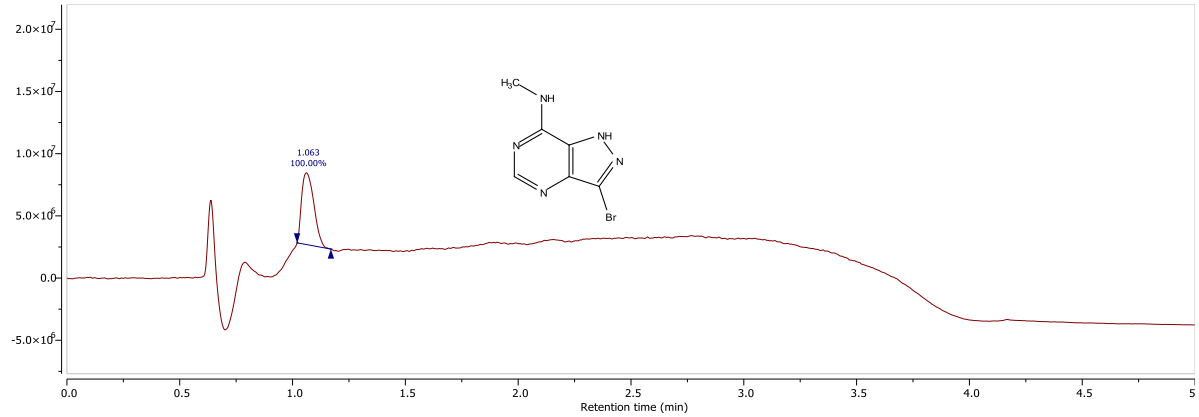


HPLC traces of final compounds

N-methyl-3-phenyl-1H-pyrazolo[4,3-d]pyrimidin-7-amine (7):



3-bromo-N-methyl-1H-pyrazolo[4,3-d]pyrimidin-7-amine (8):



Abbreviations

MD, molecular dynamics; RMSD, root-mean-square-deviation; LC-HR- ESI-MS, liquid chromatography high resolution electrospray ionization mass spectrometry; bs, broad signal.