Supplementary Data

Biochemical and structural basis for YTH domain of human YTHDC1 binding to methylated adenine in DNA

Clayton B. Woodcock[1,3], John R. Horton[1,3], Jujun Zhou[1], Mark T. Bedford[1], Robert M. Blumenthal[2], Xing Zhang[1,4], Xiaodong Cheng[1,4]

[1]Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[2]Department of Medical Microbiology and Immunology, and Program in Bioinformatics, The University of Toledo College of Medicine and Life Sciences, Toledo, OH 43614, USA

[3]These authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors: C.B.W. and J.R.H.

[4]To whom correspondence may be addressed: xzhang21@mdanderson.org; xcheng5@mdanderson.org

Email addresses of all authors:

CBW (CBWoodcock@mdanderson.org); JRH (JRHorton@mdanderson.org); JZ (JZhou12@mdanderson.org); MTB (mtbedford@mdanderson.org); RMB (robert.blumenthal@utoledo.edu); XZ (xzhang21@mdanderson.org); XC (xcheng5@mdanderson.org)

6 Figures and 2 Tables

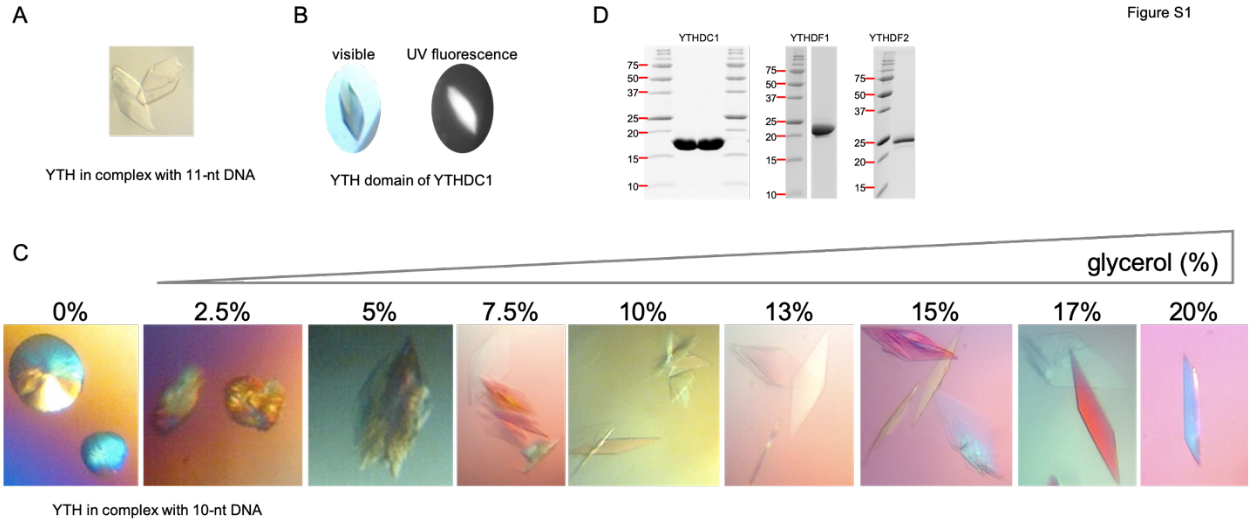


Figure S1. Crystals of YTH of YTHDC1 in complex with DNA. (**A**) Examples of crystals of YTHDC1 YTH domain in complex with 11-nt DNA. (**B**) Example of crystal of YTHDC1 YTH domain without DNA. (**C**) Crystallization of YTH-DNA (10-nt) under increased glycerol concentration. (**D**) Examples of purified YTH domains in 18% PAGE used in the study.

Figure S2

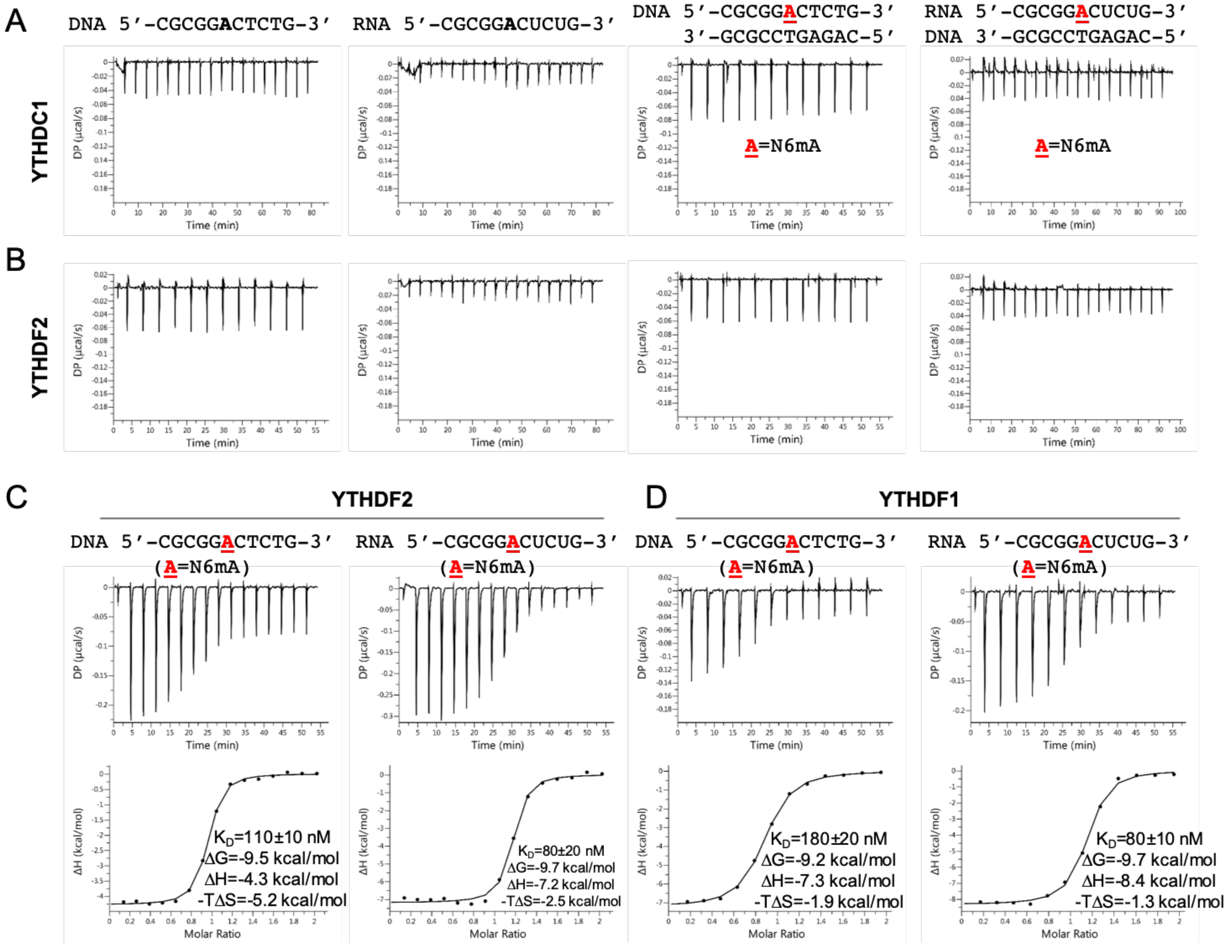


Figure S2. (**A-B**) ITC measurements of YTHDC1 (panel A) and YTHDF2 (panel B) with oligonucleotides containing no modifications or hemi-methylated double-stranded DNA or RNA/DNA hybrid as indicated on top of panels. (**C-D**) ITC measurements of YTHDF2 (panel C) and YTHDF1 (panel D) binding methylated ssDNA or methylated ssRNA.

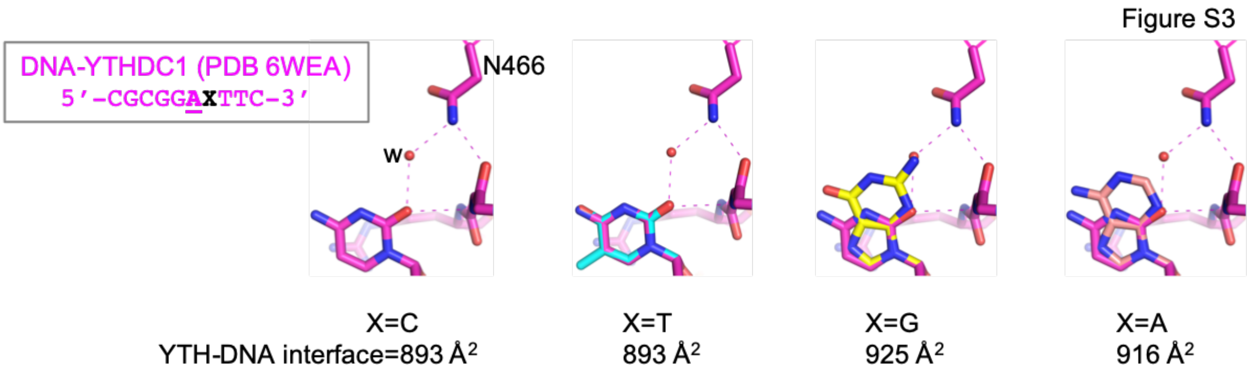


Figure S3. A model of varied nucleotide (X=C, T, G and A) immediately after N6mA, with calculated YTH-DNA interface. The protein residue Arg475 is in the background for stacking with the base. A larger purine base (G and A) would increase the interface with Arg475. The guanine *N2* atom could replace the water-mediated interaction with the side chain of Asn466.

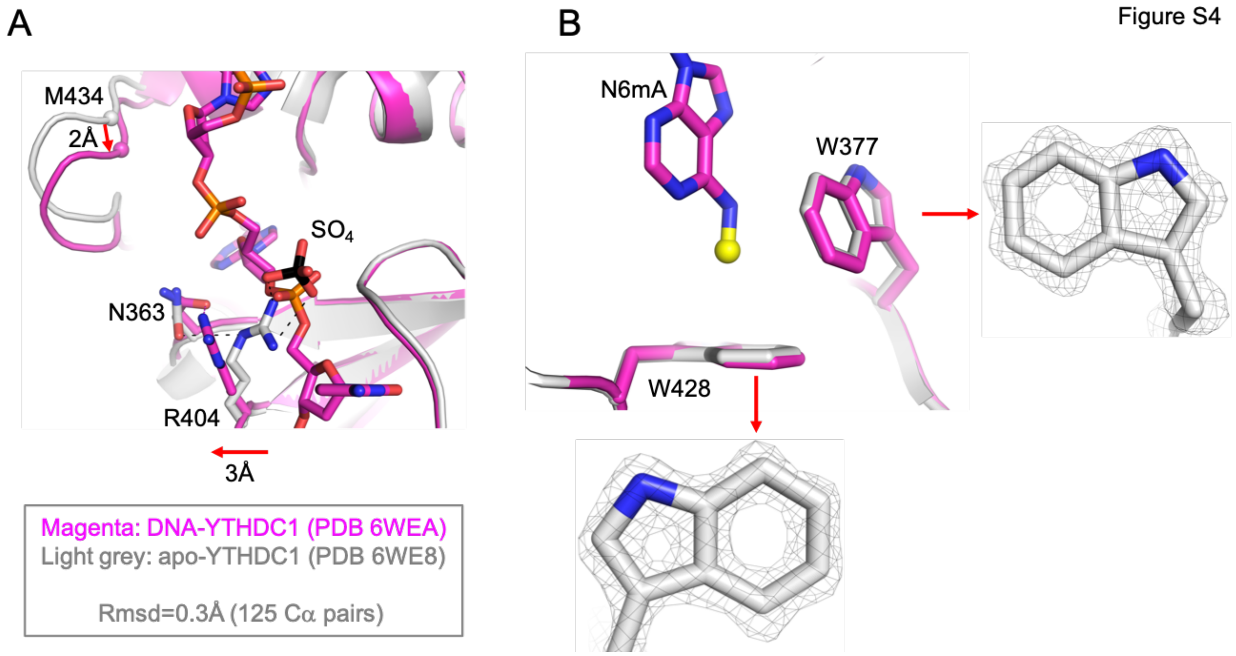


Figure S4. Comparison of YTH domain of YTHDC1 in the presence and absence of DNA. (**A**) The concurrent movement of Met434-containing loop and Arg404 upon DNA binding. (**B**) The N6mA-binding aromatic cage is rigid. The electron density (2Fo-Fc) is contoured at 2σ above the mean for Trp428 and Trp377 respectively.

Figure S5

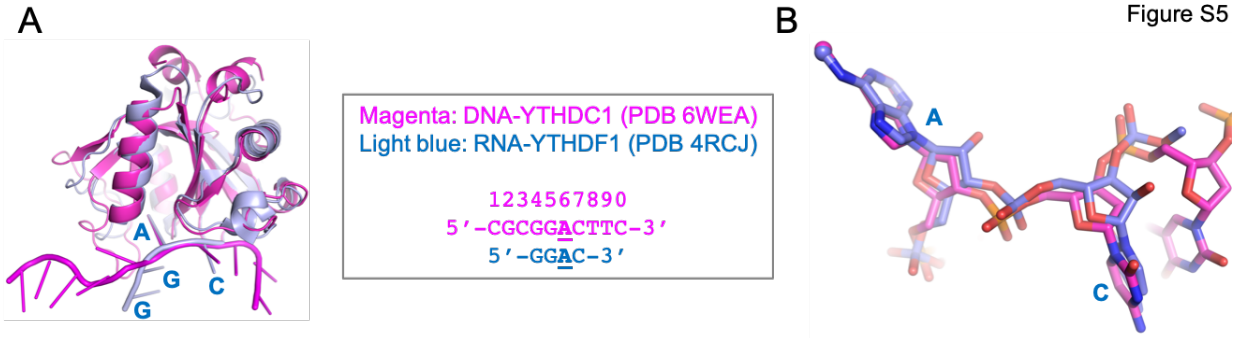
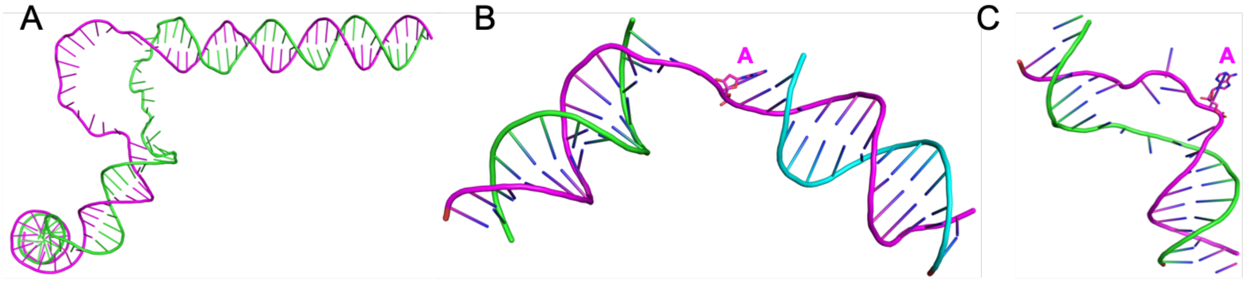


Figure S5. Comparison of YTHDC1-DNA and YTHDF1-RNA interactions. (**A**) Superimposition of the two YTH structures. Note only the N6mA-C dinucleotides are superimposable. (**B**) The cytosine base is rotated 180º along the glycosidic bond.

Figure S6



(PDB 5IYB) (PDB 5CO8) (PDB 6PBD)

Figure S6. **Examples of DNA structures deviating from non-B conformation**. (**A**) Human transcription promoter opening results in strand separation (PDB 5IYB). (**B**) The Holliday junction-resolving enzyme bound to product DNA containing a single-base gap at the bent position (PDB 5CO8). (**C**) CcrM-bound strand separation with a flipped-out adenine (PDB 6PBD).

Table S1. Summary of X-ray data collection at wavelength=1Å and refinement statistics (*)

PDB ID	6WE8	6WE9	6WEA
DNA (**A**=N6mA)	none	5'-CGCGG**A**CTCTG-3'	5'-CGCGG**A**CTTC-3'
Space group	$P2_1$	$C2$	$P2_1$
Cell dimensions (Å)	39.91, 103.96, 42.20	119.97, 42.54, 90.02	82.94, 42.64, 119.87
α, β, γ (°)	90, 104.9, 90	90, 103.5, 90	90, 92.4, 90
Resolution (Å)	20.01-1.18 (1.22-1.18)	35.31-1.59 (1.65-1.59)	40.59-1.79 (1.85-1.79)
[a] R_{merge}	0.072 (0.579)	0.204 (0.693)	0.134 (0.889)
R_{pim}	0.038 (0.517)	0.094 (0.543)	0.083 (0.678)
$CC_{1/2}$, CC	(0.410, 0.763)	(0.468, 0.799)	(0.645, 0.886)
[b] <I/σI>	22.2 (1.2)	10.4 (1.9)	8.7 (1.1)
Completeness (%)	86.1 (see below) 1.33-1.27 (89.2) 1.27-1.22 (55.6) 1.22-1.18 (26.5)	91.6 (63.6)	96.9 (84.3)
Redundancy	5.8 (1.4)	4.7 (1.7)	3.4 (2.2)
Observed reflections	536,287	253,407	251,926
Unique reflections	93,137 (2,859)	54,367 (3,725)	73,744 (6,380)
Refinement			
Resolution (Å)	1.18	1.59	1.79
No. reflections	93,024	54,294	73,624
[c] R_{work} / [d] R_{free}	0.167 / 0.195	0.222 / 0.246	0.210 / 0.237
No. Atoms (ASU)	2 proteins	2 complexes	4 complexes
Protein	2642	2523	5170
DNA	-	390	804
Solvent	454	228	612
B Factors (Å^2)		Mol A / Mol B [e]	
Protein	17.8	25.4 / 110.0	26.7
DNA	-	48.1 / 94.8	23.2
Solvent	31.6	38.6	27.2
R.m.s. deviations			
Bond lengths (Å)	0.002	0.003	0.004
Bond angles (°)	0.6	0.6	0.7

* Values in parenthesis correspond to highest resolution shell.

[a] R_{merge}=Σ|I-<I>|/ΣI, where I is the observed intensity and <I> is the averaged intensity from multiple observations.

[b] <I/σI>=averaged ratio of the intensity (I) to the error of the intensity (σI).

[c] R_{work}=Σ|Fobs-Fcal|/Σ|Fobs|, where Fobs and Fcal are the observed and calculated structure factors, respectively.

[d] R_{free} was calculated using a randomly chosen subset (5%) of the reflections not used in refinement.

[e] One complex (B) of two in the asymmetric unit has much weaker electron density and averages of each are given.

Table S2. Comparison of three YTH-Domain Protein Orthologous Groups Across Vertebrate Classes

(cyan or white on black=invariant in all three protein families; white letter on red=residues interaction with N6mA and the 3’ cytosine)

```
                                                  361/2            377
YTHDC1
Mammalia       345                  TSKLKYVLQDARFFLIKSNNHENVSLAKAKGVWSTLPVNEKKLNLAFRSA---RSVILIFSVR
Aves           322                  TSKgKYILQDARFFLIKSNNHENVSLAKAKGVWSTLPVNEKKLNAAFRSA---RSVILIFSVR
Reptilia       311                  TSRLKYILQDARFFLIKSNNHENVSLAKAKGVWSTLPVNEKKLNAAFRSA---RSVILIFSVR
Amphibia       325                  TNKIRYILQEARFFLIKSNNHENVSLAKAKGVWSTLPVNEKKLNAAFRSA---RSVILVFSVR
Osteichthyes   342                  TSKLKYILRDARFFLIKSNNHENVSLAKAKGVWSTLPVNEKKLNAAFRSA---RSVILIFSVR
Chondrichthyes 361                  TSKLKYILRDARFFLIKSNNHENVSLAKAKGVWSTLPVNEKKLNTAFRDA---RSVILIFSVR
YTHDF1
Mammalia       332 VESHPVLEKLKAAHSYNPKEFEWNLKSGRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRCMSSKGPVYLLFSVN
Aves           334 GESHPVLEKLKAAHSYNPKDFEWNLKNGRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRSMNSKGPVYLLFSVN
Reptilia       332 GESHPVLEKLKAAHSYNPKDFEWNLKNGRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRSMNSKGPVYLLFSVN
Amphibia       332 NESHPVLEKLKAAHSYNPKDFDWNLKNGRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRSMNGKGPVYLLFSVN
Osteichthyes   339 AESHPVLEKLRAAHSYNPKDFDWNVKNARVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRAMNGKGPVYLLFSVN
Chondrichthyes 333 NETHPVLEKLKAAYCYNPKDFDWNLKNGRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAYRSMAGKAPVYLLFSVN
YTHDF2
Mammalia       391         KLRSINNYNPKDFDWNLKHGRVFIIKSYSEDDIHRSIKYNIWCSTEHGNKRLDAAYRSMNGKGPVYLLFSVN
Aves           393         KLRSINNYNPKDFDWNPKHGRVFIIKSYSEDDIHRSIKYNIWCSTEHGNKRLDAAYRSMNGKGPVYLLFSVN
Reptilia       392         KLRSINNYNPKDFDWNPKHGRVFIIKSYSEDDIHRSIKYNIWCSTEHGNKRLDAAYRSMNGKGPVYLLFSVN
Amphibia       412         KLRSVNNYNPKDFDWNPKHGRVFIIKSYSEDDIHRSIKYNIWCSTEHGNKRLDAAYRSLAAKGPLYLLFSVN
Osteichthyes   335         KLRSVNNYNPKDFDWNPKHGRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDAAYRTMNGKGPVYLLFSVN
Chondrichthyes 376         KLKAAHSYNPKDFDWNLKNGRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRSMNGKGPVYLLFSVN

                                         428                                   466      475/6
YTHDC1
Mammalia           ESGKFQGFARLSSESHHGGSPIHWVLPAGMSAKMLGGVFKIDWICRRELPFTKSAHLTNPWNEHKPVKIGRDGQEIELECGTQL
Aves               ESGKFQGFARLSSESHHGGSPIHWVLPAGMNAKMLGGVFKIDWICRRELPFTKSAHLTNPWNEHKPVKIGRDGQEIEPECGTQL
Reptilia           ESGKFQGFARLASESHHGGSPIHWVLPAGMNARMLGGVFKIDWICRRELPFTKSSHLTNPWNEHKPVKIGRDGQEIEIECGTQL
Amphibia           ESGKFQGFARLSSESHHGGSPIHWVLPAGMNAKMLGGVFKIDWICRRELPFTKCVHLTNPWNEHKPVKIGRDGQEIEPDCGTQL
Osteichthyes       ESGKFQGFARLSSESHHGGSPIHWVLPAGMNAKMLGGVFRIDWICRRELPFTKTAHLSNPWNEHKPVKIGRDGQEIEPECGTQL
Chondrichthyes     ESGKFQGFARLSSESHHGGSPIHWVLPAGMNAKMLGGVFKIDWICRRELPFTKTGHLVNLWNENKPVKIGRDGQEIEPECGTQL
YTHDF1
Mammalia           GSGHFCGVAEMKSPVDYGTSAGVWSQD------KWKGKFDVQWIFVKDVPNNQLRHIRLENNDNKPVTNSRDTQEVPLEKAKQV
Aves               GSGHFCGVAEMKSPVDYGTSAGVWSQD------KWKGKFDVKWIFVKDVPNNQLRHIRLENNDNKPVTNSRDTQEVPLEKAKQV
Reptilia           GSGHFCGVAEMKSSVDYGTSAGVWSQD------KWKGKFDVKWIFVKDVPNNQLRHIRLENNDNKPVTNSRDTQEVPLEKAKQV
Amphibia           GSGHFCGVAEMKSPVDYGTSAGVWSQD------KWKGKFDVKWLFVKDVPNNQLRHIRLENNDNKPVTNSRDTQEVPLEKAKLV
Osteichthyes       GSGHFCGVAEMRSPVDYGTSAGVWSQD------KWKGKFDVKWLFVKDVPNSQLRHIRLENNDNKPVTNSRDTQEVPLEKAKQV
Chondrichthyes     GSGHFCGVAEMKSPVDYGTCAGVWSQD------KWKGKFDVKWIFVKDVPNNQLRHIRLENNDNKPVTNSRDTQEVPLEKAKQV
YTHDF2
Mammalia           GSGHFCGVAEMKSAVDYNTCAGVWSQD------KWKGRFDVRWIFVKDVPNSQLRHIRLENNENKPVTNSRDTQEVPLEKAKQV
Aves               GSGHFCGVAEMKSAVDYNTCAGVWSQD------KWKGRFDVRWIFVKDVPNSQLRHIRLENNENKPVTNSRDTQEVPLEKAKQV
Reptilia           GSGHFCGVAEMKSAVDYNTCAGVWSQD------KWKGRFDVRWIFVKDVPNSQLRHIRLENNENKPVTNSRDTQEVPLEKAKQV
Amphibia           GSGHFCGVAEMRSPVDYNTCAGVWSQD------KWKGRFDVRWIFVKDVPNSQLRHIRLENNENKPVTNSRDTQEVPLDKARQV
Osteichthyes       GSGHFCGVAEMKSTVDYNTCAGVWSQD------KWKGRFDVRWIFVKDVPNSQLRHIRLENNENKPVTNSRDTQEVPLEKAKQV
Chondrichthyes     GSGHFCGVAEMKSPVDYGTSAGVWSQD------KWKGKFDVKWLFVKDVPNNQLRHIRLENNDNKPVTNSRDTQEVPLEKAKLV

YTHDC1                                                          GenBank #         Species
Mammalia           CLLFPPDESIDLYQVIHKMRH                    509 NP_001026902.1 Homo sapiens
Aves               CLLFPPDESIDLYQLIHKMRH                    486 XP_025005419.1 Gallus gallus
Reptilia           CLLFPPDESIDLYQLIHKMRH                    495 XP_003222405.1 Anolis carolinensis
Amphibia           CLLFPADDSIDLYQVIHKMRH                    489 XP_012815913.1 Xenopus tropicalis
Osteichthyes       CLLFPPDESIDLYQVIHKMRH                    506 XP_015218249.1 Lepisosteus oculatus
Chondrichthyes     CLLFPPDESIDLYQFIHKMRH                    527 XP_020370216.1 Rhincodon typus
YTHDF1
Mammalia           LKIISSYKHTTSIFDDFAHYEKRQEEEEVVRKERQSRNKQ 559 NP_060268.2
Aves               LKIIATYKHTTSIFDDFSHYEKRQEEEEVVRKVNLLKNLF 568 XP_015151870.2
Reptilia           LKIIATYKHTTSIFDDFSHYEKRQEEEEVVRKERQNRNKQ 560 XP_003220752.1
Amphibia           LKIIATYKHTTSIFDDFSHYEKRQEEEEVVRKVMAKPDSC 565 NP_989392.1*
Osteichthyes       LKIIATYKHTTSIFDDFSHYEKRQEEEEVVRKNFEPPLIQ 565 XP_006639659.1
Chondrichthyes     LKIIATYKHTTSIFDDFSHYEKRQEEEEVVRKDKISLPVA 560 XP_020386177.1
YTHDF2
Mammalia           LKIIASYKHTTSIFDDFSHYEKRQEEEESVKKERQGRGK  579 NP_001166599.1
Aves               LKIIATYKHTTSIFDDFSHYEKRQEEEENVKKERQGRVK  581 XP_015153209.1
Reptilia           LKIIATYKHTTSIFDDFSHYEKRQEEEETVKKERQGRVK  580 XP_003229643.1
Amphibia           LKIIASYKHTTSIFDDFSHYEKRQEEEESVKKVEVQGSD  600 NP_989392.1*
Osteichthyes       LKIIASYKHTTSIFDDFSHYEKRQEEEECVKKERQGRVK  523 XP_006631512.1
Chondrichthyes     LKIIATYKHTTSIFDDFSHYEKRQEEEEVVRKVMAKPDS  564 XP_020366751.1
```

Notes: * = same ortholog was best *X. tropicalis* hit for both YTHDF1 and YTHDF2; for *L. camtschaticum*, no match was found having expect score of ≤10.