

## Supplemental information

### Supplemental Figure Legends

#### Figure A1. Phylogram of cathepsins.

A maximum likelihood phylogeny of aspartyl and cysteine cathepsins in *T. thermophila*, together with the most highly related homologs (as judged by BLAST scores) present in a variety of organisms from the other major eukaryotic lineages. This figure represents an extended version of Figure 1B.

#### Figure A2. Primary sequence alignment of Cathepsin 4 (Cth4p) of *T. thermophila* with *H. sapiens* Cathepsin C

Conserved catalytic motifs are highlighted in red. Conserved active site residues C, H, N are marked by arrows. GenBank™ Accession Number of *Tetrahymena* CTH4 is XM\_001023356. UniProtKB/Swiss-Prot number of human cysteine protease (CTHC) is P53634.

**Table S1.** Master Primer list

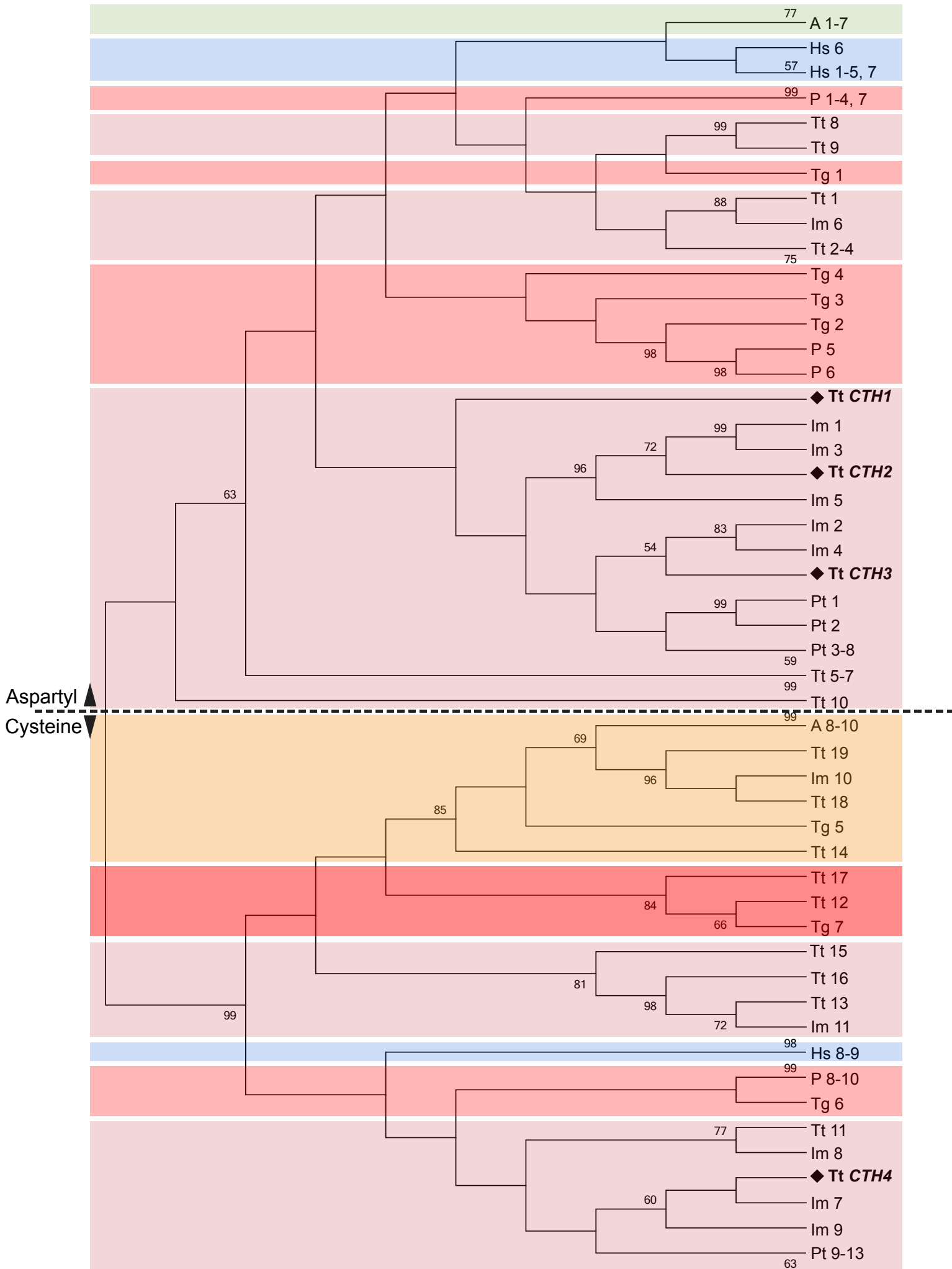
Nam	Description	5'-Sequence-3'
01A	CTH4 Forward	CACCATGGTTAAAAGTAGATTAGCTGC
01B	CTH4 Reverse	TTATTTTCTTTTAGATTTTCAACAAAGTATTTG
02A	CTH4 KO 5'UTR-Forward	GGGAACAAAAGCTGGCCATTACTTGTAACATATTATTTTATGAGGGTTTG
02B	CTH4 KO 5'UTR-Reverse	CCGCCACCGCGGTGGTCTTTCAATATTTTATTTAATTTTGTGGTTCG
03A	CTH4 KO 3'UTR-Forward	TACCGTCGACCTCGAGAAAACAACACCTGCAACATTATTAACC
03B	CTH4 KO 3'UTR-Reverse	CGGGCCCCCCTCGAGGAAAATGATAGAACTTTTGACTGATTGGG
04A	CTH4 Forward RT PCR	ATGGTAAAAGTAGATTAGCTGC
04B	CTH4 Reverse RT PCR	GAGAGCAAACAGATTTATATTATCCTG
05A	Sortilin3 Forward RT PCR	ATGCGCTAGCGGATCTGCTAAAGACGATAGCAAAGAAAGAAC
05B	Sortilin3 Reverse RT PCR	TAGAAACCATGGATCCATTTCTAGGATCAAAGGATCTTCTTAC
06A	CTH4 Forward (Endo)	ATGCGCTAGCGGATCATGGTTAAAAGTAGATTAGCTGCAAT
06B	CTH4 Reverse (Endo)	TAGAAACCATGGATCCTTATTTTCTTTTAGATTTTCAACAAAGTATTTGTC
07A	CTH4 3'UTR-Forward (Endo)	ATTCGATATCAAGCTTCGATTTCAGAAATACAAAGTTATTTTACATTTTG
07B	CTH4 3'UTR-Reverse (Endo)	CGGTATCGATAAGCTTGGAATTTGTATGAACCACTAGATTC
08A	CTH4 mutant Forward Cys352:Ala	GAAGATTGTGGAAGTGCATATGCTATTGCCACA
08B	CTH4 mutant Reverse Cys352:Ala	TGTGGCAATAGCATATGCACTTCCACAATCTTC
09A	CTH4 mutant Forward His505:Ala	TGGTAAAAGTTGATGCTTCAGTTCTCTGCTAT
09B	CTH4 mutant Reverse His505:Ala	ATAGCAGAGAACTGAAGCATCAACTTTTACCA
10A	CTH4-6His Forward	CGTGTTTAAACATGGTTAAAAGTAGATTAGCTGC
10B	CTH4-6His Reverse	TTGGGCCCTCAATGATGATGATGATGATGTTATTTTCTTTTAGATTTTCAACAAAGTAT

**Table S2.** Accession numbers from which Cathepsin sequences were obtained to create the phylogeny in Figure1B and A1

<b>Organism</b>	<b>Abbreviation</b>	<b>Accession</b>	<b>Definition</b>
<i>Toxoplasma gondii</i> ME49	Tg1	XP_002367480.1	Eukaryotic aspartyl protease, putative
<i>Toxoplasma gondii</i> ME49	Tg2	XP_002365394.1	Eukaryotic aspartyl protease, putative
<i>Toxoplasma gondii</i> ME49	Tg3	XP_002367043.1	Eukaryotic aspartyl protease, putative
<i>Toxoplasma gondii</i> ME49	Tg4	TGME49_272510	Aspartyl protease
<i>Toxoplasma gondii</i> ME49	Tg5	XP_002367300.1	Cysteine proteinase, putative
<i>Toxoplasma gondii</i> ME49	Tg6	XP_002371619.1	Cathepsin C2 (TgCPC2)
<i>Toxoplasma gondii</i> ME49	Tg7	EPT32244.1	Cathepsin CPL
<i>Plasmodium falciparum</i> 3D7	P1	XP_001348248.1	Plasmepsin IV
<i>Plasmodium falciparum</i>	P2	AAW71461.1	Plasmepsin 4, partial
<i>Plasmodium falciparum</i>	P3	AAW71459.1	Plasmepsin 4, partial
<i>Plasmodium falciparum</i>	P4	AAW71460.1	Plasmepsin 4, partial
<i>Plasmodium falciparum</i> 3D7	P5	XP_001351190.1	Plasmepsin VI
<i>Plasmodium knowlesi</i> strain H	P6	XP_002258928.1	Aspartyl protease
<i>Plasmodium falciparum</i>	P7	AAW71452.1	Plasmepsin 2, partial
<i>Plasmodium cynomolgi</i> strain B	P8	XP_004222389.1	Cathepsin C precursor
<i>Plasmodium knowlesi</i> strain H	P9	XP_002259168.1	Cathepsin c precursor
<i>Plasmodium vivax</i> Sal-1	P10	XP_001615284.1	Cathepsin C precursor
<i>Homo sapiens</i>	Hs1	AAI71897.1	Pepsinogen 5, group I (pepsinogen A)
<i>Homo sapiens</i>	Hs2	NP_055039.1	Pepsin A-5 preproprotein
<i>Homo sapiens</i>	Hs3	NP_001073276.1	Pepsin A-4 preproprotein
<i>Homo sapiens</i>	Hs4	BAF84553.1	Unnamed protein product
<i>Homo sapiens</i>	Hs5	NP_001901.1	Cathepsin E isoform a preproprotein
<i>Homo sapiens</i>	Hs6	NP_001900.1	Cathepsin D preproprotein
<i>Homo sapiens</i>	Hs7	NP_001901.1	Cathepsin E isoform a preproprotein
<i>Homo sapiens</i>	Hs8	EAW59363.1	Cathepsin C, isoform CRA_a
<i>Homo sapiens</i>	Hs9	BAG58920.1	Unnamed protein product
<i>Arabidopsis thaliana</i>	A1	NP_193936.2	Aspartyl protease family protein
<i>Arabidopsis thaliana</i>	A2	CAA18108.1	Aspartic proteinase like protein
<i>Arabidopsis thaliana</i>	A3	NP_172655.1	Aspartic proteinase A1
<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>	A4	XP_002892661.1	Aspartyl protease family protein
<i>Arabidopsis thaliana</i>	A5	NP_176419.2	Aspartic proteinase A2
<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>	A6	XP_002872693.1	Aspartyl protease family protein
<i>Arabidopsis thaliana</i>	A7	NP_192355.1	Aspartic proteinase A3
<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>	A8	XP_002889395.1	Hypothetical protein ARALYDRAFT_887368
<i>Arabidopsis thaliana</i>	A9	NP_563648.1	Putative cathepsin B-like cysteine protease
<i>Arabidopsis thaliana</i>	A10	NP_567215.1	Putative cathepsin B-like cysteine protease
<i>Tetrahymena thermophila</i>	CTH1	XP_001014915.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	CTH2	XP_001008250.2	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	CTH3	XP_001012968.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	CTH4	XP_001023356.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt1	XP_001016313.1	Eukaryotic aspartyl protease family protein (macronuclear)

<i>Tetrahymena thermophila</i>	Tt2	XP_001017733.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt3	XP_001027456.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt4	XP_001030703.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt5	XP_001026655.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt6	XP_001026656.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt7	XP_001010205.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt8	XP_001010694.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt9	XP_001010693.1	Eukaryotic aspartyl protease family protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt10	XP_001012969.1	Hypothetical protein THERM_00321690 (macronuclear)
<i>Tetrahymena thermophila</i>	Tt11 (CTH30)	XP_001022323.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt12 (CTH32)	XP_001029901.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt13	XP_001471259.1	Cathepsin z (macronuclear)
<i>Tetrahymena thermophila</i>	Tt14	XP_001015806.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt15	XP_001022043.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt16	XP_001031483.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt17	XP_001020178.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt18 (CTHB)	XP_001012594.1	Papain family cysteine protease containing protein (macronuclear)
<i>Tetrahymena thermophila</i>	Tt19	XP_001015333.1	Papain family cysteine protease containing protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt1	XP_001452128.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt2	XP_001459126.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt3	XP_001453245.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt4	XP_001454430.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt5	XP_001456226.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt6	XP_001442519.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt7	XP_001452716.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt8	XP_001438277.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt9	XP_001455775.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt10	XP_001431306.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt11	XP_001429095.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt12	XP_001459019.1	Hypothetical protein (macronuclear)
<i>Paramecium tetraurelia strain d4-2</i>	Pt13	XP_001448556.1	Hypothetical protein (macronuclear)
<i>Ichthyophthirius multifiliis</i>	Im1	EGR30663.1	WD repeat protein
<i>Ichthyophthirius multifiliis</i>	Im2	EGR30870.1	Hypothetical protein IMG5_122010
<i>Ichthyophthirius multifiliis</i>	Im3	EGR34063.1	Hypothetical protein IMG5_024830
<i>Ichthyophthirius multifiliis</i>	Im4	EGR29114.1	Hypothetical protein IMG5_162980
<i>Ichthyophthirius multifiliis</i>	Im5	EGR31334.1	Hypothetical protein IMG5_112710
<i>Ichthyophthirius multifiliis</i>	Im6	EGR32788.1	Hypothetical protein IMG5_070700, partial
<i>Ichthyophthirius multifiliis</i>	Im7	EGR30116.1	Hypothetical protein IMG5_141560
<i>Ichthyophthirius multifiliis</i>	Im8	EGR34799.1	Hypothetical protein IMG5_001760
<i>Ichthyophthirius multifiliis</i>	Im9	EGR34889.1	Nucleotide binding protein, putative
<i>Ichthyophthirius multifiliis</i>	Im10	EGR28345.1	Hypothetical protein IMG5_177790
<i>Ichthyophthirius multifiliis</i>	Im11	EGR34021.1	Papain family cysteine protease, putative





Color Key: ■ Plants ■ Animals ■ Ciliates ■ Apicomplexans ■ Ciliates and Apicomplexans ■ Ciliates, Apicomplexans, and Plants

```

hCTHC MGAGPSLLLAALLLLLSDGDVAVRCDTPANCTYLDLLGTWVFQVG---SSGSQRDVCN--- 54
TtCTH4 --MVKSRLAAILLLEFAVQSLLVQADLPVHCVRSQITGKWLLEKQKVLASSNNKPVSCGHQ 58
      * * * ** :    * : * * * : * : : * : * : : * : * : * : *
hCTHC -----SVMGPOEKKVVVYLQKLDATYDDLGNLSGHTIIYNQGFIEVLNDYKW 101
TtCTH4 IPDKAESSNSAMGGVLTVEVEYEYVFLDEDFTVRGKEGK-GRWTLVYDEGWEIEYNGVKY 117
      . * :    * : : * : * : : * . . . * : * : * : * : * : * * * . * :
hCTHC FAFFKYKEEGS-KVTTYCNETMTGWVHDLGRNWACFTGKVKGTASEN----- 148
TtCTH4 THFFKYAKDDSGQYKSVCSQTLVGWFNHIQKRSRGSQAIKADASQEFSDNVQQAHVVQP 177
      **** : : * : : : * : * : * : * : : : : . . * . * . : * : *
hCTHC -----VYVNI AHLKNSQE-----KYSNRLYKYDHN---- 173
TtCTH4 KETQYIFSEISSTMNKLSSQEQQKSKINASNLSLKRQQLSSSKNKFHQSIYLPHENVFA 237
      :    : : * . . * *          * : : : *    * :
hCTHC -----FVKAINAIQKS-WTATTYMEYETLTLGDMIRRSKGHSRKIIPRPKA 218
TtCTH4 QMKQTPATLLNHQEIVKRLNSIKGSTWQAKVPSFISKMSLEQMNTTRAGLRMDKPKFSYDN 297
      : * * : * * : * * * . . : : * : * * * : * ▼ .
hCTHC PLTAEIQQKILHLP-----TSWDWRNVHGIN-----FVSPVRNQASCGSCYSFAS 263
TtCTH4 KQGSFIQKSSSSLRGNKATTDYAHFSMENISDLPKSFDAWAEYIHAPRSQEDCGSCYAIAAT 357
      : * * : .    *        * : :    . . . * .          : : . * . * . * : : * :
hCTHC MGMLEARIRILTNNSQTPILSPQEVVSCSQYAQGCCEGGFPYLIAGKYAQDFGLVEEACFP 323
TtCTH4 TSMLSSRLWIKYGDN--TQLSPQHSLACNYNQCGDGGYGLVLS-KFYSEFEAVPESCHP 414
      . * * : * : * . . . * * * . : * . * * * * : * : * : * . * * * * .
hCTHC YTGTDS---PCKMKEDCFRYYSSEYHYVGGFYGGCNEALMKLELVHGHGPMVAFAFEVYDDF 380
TtCTH4 YEARDGQC�HCNVE SLSEVFTVTDYEYIGGSYKSTERLMMEIEYKNGPIVVSFEPKMF 474
      * . * .      * : : . . :    : * : * * ▼ * * . . * * * : : * : * * ▼ * * **
hCTHC LHYKKGIIYHHTGLRDPFN-----PFELTNHAVLLVGYGTDSASGMDYIWIKNSWGTGWG 434
TtCTH4 MYYNKGIYHSVDANQWIQNNNEENPVWQKVDHSVLCYGWGEDENG--KFWLLQNSWGEEWG 532
      : * : * * * * . . : : : : : : * : * * * * : * * . . : * : : * * * * **
hCTHC ENGYFRIRRGTTDECAIESIAVAATPIPKL----- 463
TtCTH4 ENGNFRMRRTDESNIEMGERANIVKTARKSPNTTEFSSTYSSHSKYFVEKSKRKQ 590
      *** * : * * * * . * * : .    * . :

```