BMC Microbiology

Research article

Biochemical and molecular characterisation of *Tetrahymena thermophila* extracellular cysteine proteases

Lutz Herrmann[†1], Michael Erkelenz[†2], Ingo Aldag[1], Arno Tiedtke[2] and Marcus WW Hartmann*[1]

Address: [1]Cilian AG, Johann-Krane-Weg 42, Muenster D-48149, Germany and [2]Institute for General Zoology and Genetics, University of Muenster, Schlossplatz 5, Muenster D-48149, Germany

Email: Lutz Herrmann - herrmann@cilian.de; Michael Erkelenz - erkelenz@uni-muenster.de; Ingo Aldag - aldag@cilian.de; Arno Tiedtke - tiedtke@uni-muenster.de; Marcus WW Hartmann* - hartman@cilian.de

* Corresponding author   †Equal contributors

Abstract

Background: Over the last decades molecular biologic techniques have been developed to alter the genome and proteome of *Tetrahymena thermophila* thereby providing the basis for recombinant protein expression including functional human enzymes. The biotechnological potential of *Tetrahymena* has been proved in numerous publications, demonstrating fast growth, high biomass, fermentation in ordinary bacterial/yeast equipment, up-scalability, existence of cheap and chemical defined media. For these reasons *Tetrahymena* offers promising opportunities for the development of a high expression system. Yet optimised high yield strains with protease deficiency such as commonly used in yeast and bacterial systems are not available.

Results: This work presents the molecular identification of predominant proteases secreted into the medium by *Tetrahymena thermophila*. A one-step purification of the proteolytic enzymes is described.

Conclusion: The information provided will allow silencing of protease activity by either knock out methods or by *Tetrahymena* specific antisense-ribosome-techniques. This will facilitate the next step in the advancement of this exciting organism for recombinant protein production.

Background

Tetrahymena thermophila is one of the most extensively examined ciliated protozoa and for decades it has served as a model organism in different research areas. The discovery of telomeres [1] and telomerase [2] as well as RNA-mediated catalysis [3] were discovered and studied in *Tetrahymena*. Within the last decades molecular biological techniques have been developed to alter *T. thermophila*'s genome and proteome: There are DNA transfection methods that allow transformation of either the germline micronucleus (MIC) or the vegetative macronucleus (MAC) [4-6]. Episomal plasmids based on an rDNA-replicon are available [7]. Homologous recombination in either MIC or MAC enable knock-out/-in techniques [8,9]. On protein level heterologous expression of related species has been performed [10,11] and recently the expression of functional human enzymes including proper formation of disulfide bridges and addition of N-glycans has been demonstrated (submitted, BMC Biotechnol). In parallel a few research groups have evaluated the

biotechnological potential of *Tetrahymena* within the last years [12-14]. Promising results have been achieved to culture this species with fast growth to high biomasses. Furthermore fermentation processes can be performed in ordinary bacterial/yeast equipment. Additionally up-scalability, one of the most important criteria for industrial production, has been demonstrated successfully. There are also data on *Tetrahymena* continuous high-cell-density fermentation in a perfused bioreactor making this organism even more useful for industrial applications [13]. For all these reasons *Tetrahymena thermophila* has been selected by the US National Human Genome Research Institute as one of the high-priority genomes for sequencing in 2002. Today sequencing of the MAC genome has nearly been completed by The Institute for Genomic Research thus enabling mining of the organism's genome. Tiny amounts of protein, e.g., can be characterized easily by mass spectrometry in connection with a search for obtained peptide fragments within the database. *T. thermophila* secretes many lysosomal enzymes into the surrounding medium [15]: phospholipases [16-18], glycosidases [19], phosphatases [20] and proteases [21] that will modify or even degrade potential product. In a heterologous expression system these undesired enzyme activities must be depressed to assure quality and yield of the product. Today all microbial expression systems can rely on decades of research results including detailed information on the genome and proteome of the used organisms. So, e.g., in *E. coli* systems optimised strains with additional tRNAs and/or protease deficiencies have been engineered and have been available on the market for many years. Yet *Tetrahymena's* commercial potential has not been exploited at all although today all necessary tools for genetic engineering are available. Till now only one protease-sequence for *Tetrahymena* has been identified by experimental means: Tetrain in *T.pyriformis*, a cathepsin L family member [22]. In *T.thermophila* a cDNA encoding for a similar putative protease, pCyp, has been described, but examination of the protein and its expression are lacking [23]. Although Straus *et al.* described the purification of different *T.thermophila* cysteine proteases by conventional chromatography methods they were not able to determine the sequence of the enzymes [21]. As all proteases described so far are cysteine proteases and nearly all proteolytic activity of cell extracts can be inhibited by cysteine protease specific inhibitors we chose a straightforward one-step purification approach described by Greenbaum *et al.* This method makes use of modified trans-epoxysuccinyl-L-leucylamido-(4-guanidino)-butane (E64) and has been adopted successfully for cysteine proteases of different organisms ranging from plants [24] over *P. falciparum* [25] to *D.melanogaster* [26]. Here we provide an additional step in the development of a high performance expression system based on *Tetrahy-*
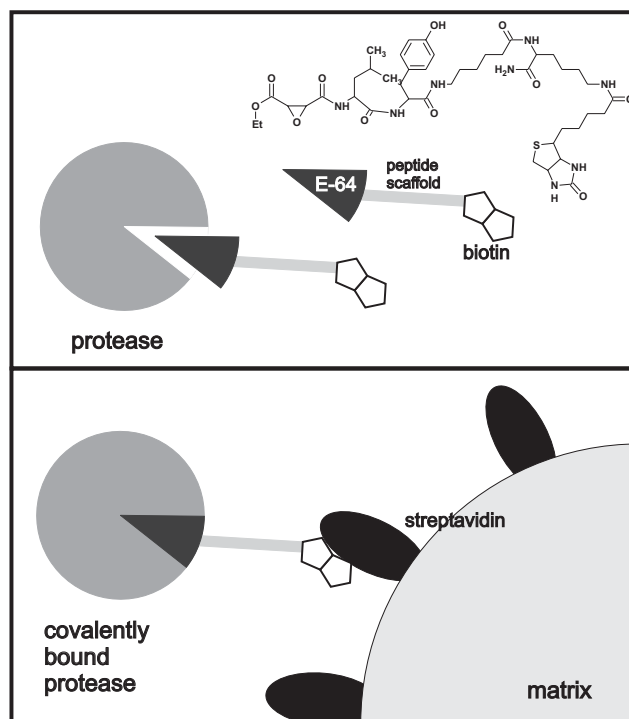

Figure 1
Purification scheme DCG-04. The structure of DCG-04 is shown in the upper part. The E64-component binds covalently to active cysteine proteases. Subsequently protease/inhibitor complex is immobilized by a streptavidin bead via the biotin tag. (lower part).

mena by an in-detail molecular characterization of the major extracellular proteases.

Results
Synthesis of DCG-04
Protease activity profiling is based on labelled protease inhibitors that covalently bind to proteases in an activity dependent manner. These specifically binding reagents can also be used for purification purposes. By linking the cysteine protease inhibitor E64 to biotin Greenbaum *et al.* created a versatile tool called DCG-04 for rapid purification of cysteine proteases by immobilizing the substrate/protease complex to streptavidin beads. The method is summarized in figure 1. Synthesis of the substance was performed according to the methods section. The formation of the proper product was verified by mass spectrometry (data not shown).

Secretion kinetics, production and ex ante characterization of **T.** thermophila *extracellular proteases*
For production of extracellular proteases in growing *T. thermophila* the secretion kinetics of the enzymes during

Figure 2
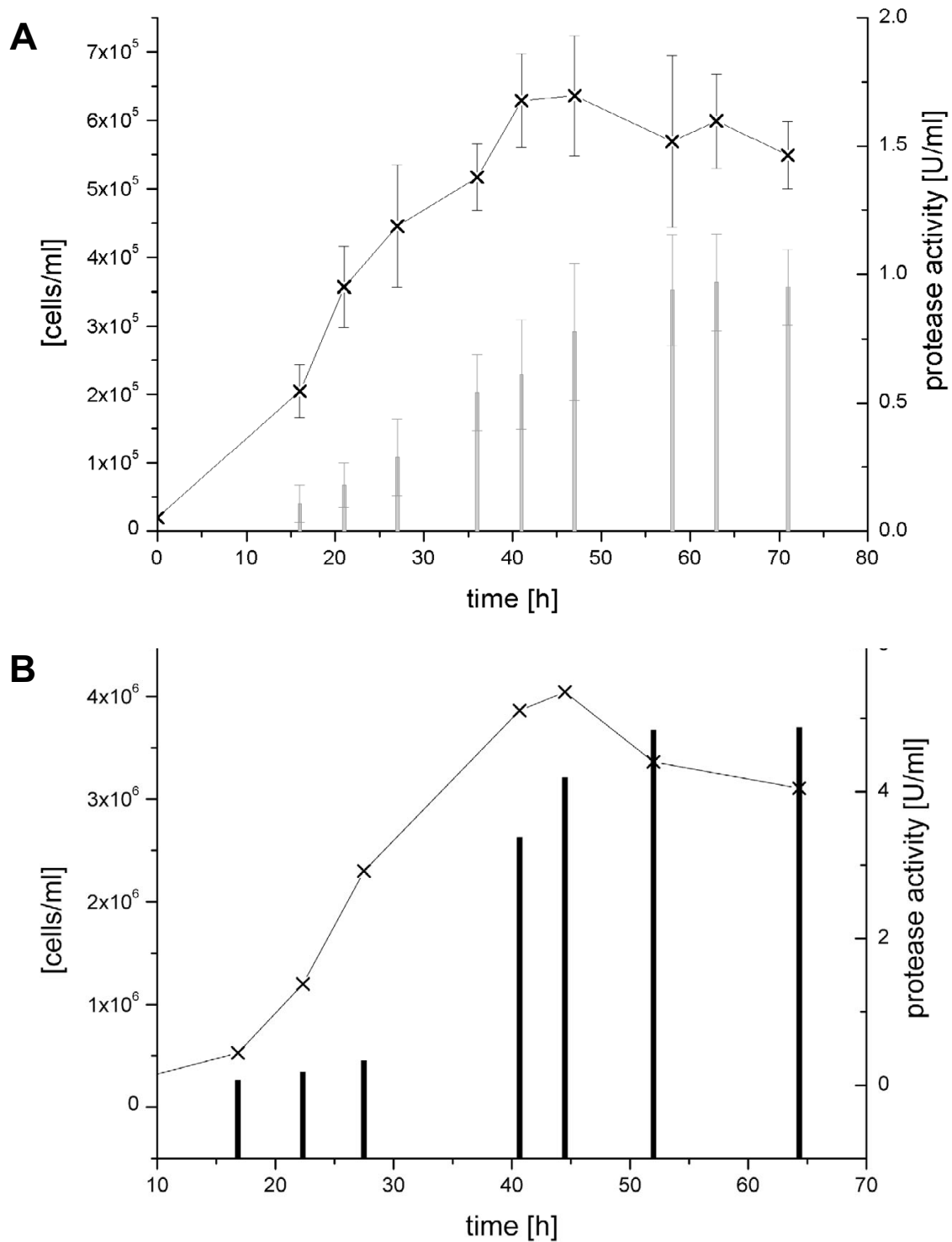**Growth curve and secretion kinetics of proteases.** A shows the mean cell titers (curve with crosses) and protease activities (bars) during shaker flask cultivation of *T.thermophila* cells. Error bars represent standard errors of three independent experiments. B depicts the cell titers (curve with crosses) and protease activity (bars) during the fermentation used for identification of extracellular proteases are shown.
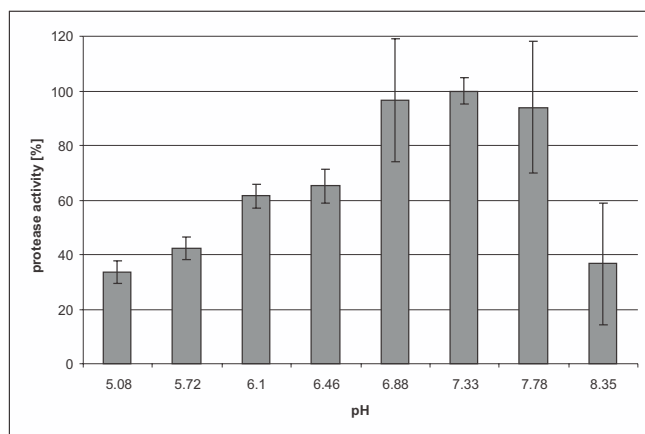
Figure 3
pH dependency of protease activities. The pH profile of protease activity in *T. thermophila* supernatants is shown. The overall activity has its maximum at neutral to slightly basic pH. Error bars represent standard errors of three independent experiments.

cultivation in shaker flasks was monitored (figure 2A). Maximum protease activity is observed in late logarithmic growing cell cultures. Consequently to achieve large amount of material, cells were separated from the media of a 2L fermentation process after 65 hours yielding 5 U/ml of protease activity (figure 2B).

To evaluate the optimal pH at which the purification would work best the protease activity in the harvested supernatant was determined at different pH values. The results shown in figure 3 suggest optimal conditions at neutral to slight basic pH.

To demonstrate that most of the proteolytic activity is due to cysteine and not serine or threonine proteases the inhibitory effect of DCG-04 on the concentrated supernatant was investigated. The strongly alkylating agent lithium iodoacetate served as positive control for enzyme inhibition. Nearly all proteolytic activity vanished by addition of DCG-04 (figure 4); remaining activity is as low as background activity. These findings argue that most of the predominant, secreted proteases are members of the cysteine protease family.

Purification and identification of T. thermophila *secreted proteases*

According to the pH activity profile of secreted proteases of *Tetrahymena* the capturing step with DCG-04 was performed at pH 7.4. Figure 5 illustrates the results of the one-step purification process: The crude supernatants before and after incubation with DCG-04 and streptavidin labelled beads (lane 1 and 2) show a vast and complex band pattern of different sized proteins. The purified pro-

tein fraction eluted from the matrix results in predominant bands running at molecular weights ranging from 22 to 28 kD. These sizes have been described for many mature cysteine proteases. To verify the specificity of the purification process, aliquots of the samples were separated by 2-D gel electrophoresis, then blotted to nitrocellulose and finally probed with an anti-biotin antibody. All spots visible on a silver stained reference gel were readily detected by the antibody (data not shown). This argues for an efficient and covalent binding of DCG-04 to the proteases. The bands between 20 and 30 kD were excised and subjected to mass spectrometry. The search algorithm exploited a preliminary database provided by TIGR, a very useful and valuable tool the importance of which has already been predicted in 2000. [27] Six different cysteine proteases of the cathepsin family were unambiguously identified by at least two independent peptides (highlighted in figure 6) termed TtCysP1-6 (*Tetrahymena thermophila* cysteine proteases 1-6). These proteins are listed in the preliminary gene release at The Institute for Genomic Research under temporary identifier 67.m00244, 103.m00134, 103.m00129, 123.m00091, 24.m00272 and 125.m00080 respectively and are located on genomic contig SB210 8254459, SB210 8253891, SB210 8253891, SB210 8254367, SB210 8254649 and SB210 8254370. Whereas the protease pCyp described by Karrer *et al.* [23] was not detectable, sequence alignments reveal that TtCysP6 is *T. thermophila* Tetrain [22] is 72 % identical to the *T.pyriformis* protein. All identified enzymes show a prepro-peptide structure indicated by an ERFNIN motif [23] (figure 6). Signal peptides were predicted by computational analysis (SignalP algorithm [28], table 1). All amino acids essential for enzyme activity are highly conserved (figure 6). As expected no fragments within the first 180 amino acids were found because the prepro-peptide (~140 aa) is cleaved off during the processing of the enzyme. Peptides containing the catalytic cysteine at position ~160 are masked by covalently bound DCG-04.

Discussion

It is known that the majority of intra- and extracellular proteases in *Tetrahymena* are cysteine proteases. Straus et al. reported, that at least four different proteases are present in *T.thermophila* supernatants. But they were not able to derive any sequences from their data [21]. The only sequence information available on an active protease in growing *Tetrahymena* cells is Tetrain, a *T. pyriformis* derived enzyme. The results presented in this work confirm all data available so far: Chemical targeting of cysteine proteases by means of a mechanism based probe combined with mass spectrometry has allowed identification of six extracellular *T. thermophila* cysteine proteases and their sequences have been determined; the existence of a *T. thermophila* Tetrain has been confirmed by
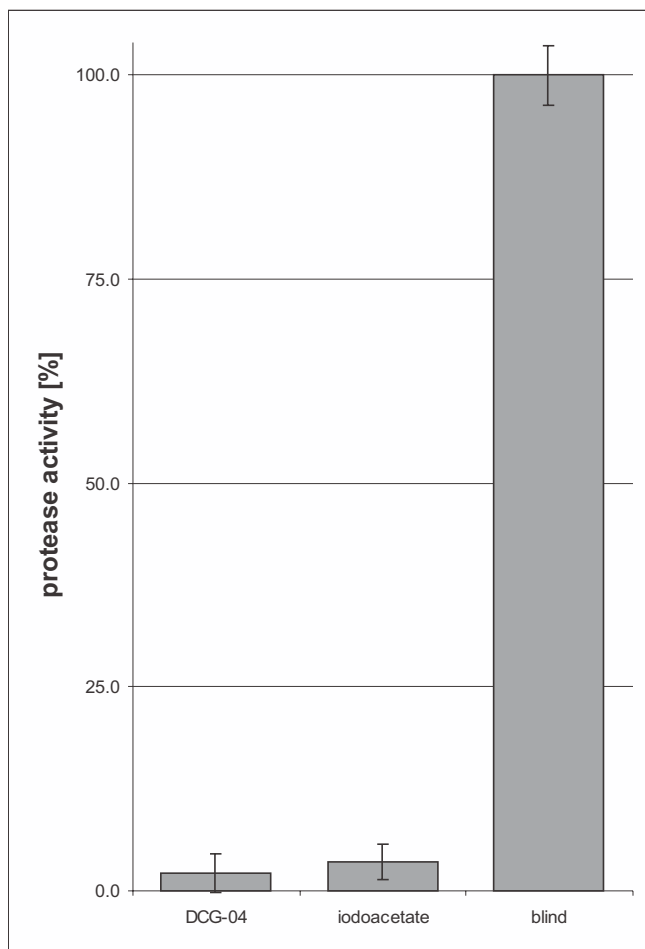
Figure 4
Inhibition of protease activity by DCG-04. Nearly all protease activity is inhibited by the addition of DCG-04 (first bar) compared to blind probe (water, very right bar). The inhibition by DCG-04 is comparable to all alkylated enzymes by iodoacetate (middle bar). Error bars represent standard errors of three independent experiments.

Figure 5
Purification of DCG-04. A silver stained SDS-gel of different purification steps by DCG-04. Lane 1 shows the crude *T. thermophila* supernatant, lane 2 the supernatant after the addition of DCG-04 and subsequent extraction by means of streptavidin beads. In lane 3 the supernatant of beads that have been washed and boiled in SDS buffer is shown. The enriched protease fraction mainly consists of bands of approximately 22-28 kD.

sequence comparison. pCyp, a protease described by Karrer [23], could not be detected in our experiments. A possible explanation is that the cDNA isolates used for detection of pCyp were derived from starved, not from growing cells. This would imply that *Tetrahymena* is able to up- and down regulate different proteases on changing environmental, physiological conditions, a phenomenon that Suzuki *et al.* already postulated for the *T. pyriformis* Tetrain protease [22]. The regulation of various proteins in *Tetrahymena* during different development stages has been reported and investigated in detail by comparing growing, starving and conjugating cells: It was shown that for many differentially expressed proteins the transcriptional activity is the major regulating mechanism[29] This could also be true for pCyp being down-regulated during

vegetative growth. Taking a look at the recently acquired genome database of *T.thermophila* one will find far more than 30 different putative genes encoding for cysteine proteases. Table 2 lists the 30 cysteine proteases that are most similar to TtCysP1. All of them have a functional signal peptide according to SignalP analysis and the conserved amino acids of the ERFNIN motif are also present in nearly any putative enzyme. These findings argue that *T.thermophila* is able to choose in a regulated way from a whole set of different proteolytic enzymes, that must be secreted. Further experiments to verify this hypothesis need to be done.

Table 1: Signal peptide prediction by SignalP analysis

Protease	TIGR temporary identifier	Signal peptide propability [%]	Max cleavage site probability [%]	Behind amino acid
TtCyp1	67.m00244	93	27	19
TtCyp2	103.m00134	98	42	21
TtCyp3	103.m00129	99	77	22
TtCyp4	123.m00091	91	38	22
TtCyp5	24.m00272	99	78	19
TtCyp6	125.m00080	89	30	20

Conclusion

The main aim in this study was to identify most active proteases in growing ciliate cells as this is the phase ideally suited for expression of foreign genes. Well established *E. coli* and yeast based expression systems have been making use of protease deficient strains for decades to enlarge their product yields. The information provided by the above results therefore is urgently needed for genetic engineering in strain optimisation. To develop a competitive, alternative expression platform based on *T. thermophila* the identified proteases must selectively be knocked out.

Methods

Synthesis of DCG-04

The active inhibitory epoxide group ethyl-(2S,3S)-oxirane-2,3-dicarboxylate was generated by adding 5.25 ml 0.1 M KOH in EtOH drop wise to 100 mg diethyl-(2S,3S)-(+)-2,3-epoxysuccinate (Acros Organics) in 3 ml EtOH on ice within two hours and further stirring at room temperature for one hour. A white precipitate formed upon standing at -20°C overnight which was recovered by evaporating EtOH. The residue was taken up by 4 ml 50 mM $NaHCO_3$ (aq) and extracted twice with 4 ml EtOAc. The aqueous phase was shifted to pH 2.5 by addition of 2 M $KHSO_4$ and was extracted five times with 3 ml EtOAc. Combined organics were dried and concentrated yielding a greenish oil which solidified upon further standing. Proper formation of product was confirmed by mass-spectrometry. Solid phase synthesis of DCG-04 was performed according to a modified protocol published by Greenbaum *et al*. Resin and amino acids were purchased from Nova Biochem. All procedures were carried out under Argon atmosphere at room temperature. Each Fmoc deprotection was carried out by two incubations (1 min and 30 min each) of 20 % piperidine in DMF. During coupling equimolar concentrated solutions of 1-hydroxybenzotriazole, N, N'-diisopropylcarbodiimide and amino acid in DMF were used. Thorough washing with DMF and CH_2Cl_2 was performed between each synthesis step. 288 mg of rink amide AM resin were swollen in DMF and deprotected. Fmoc-Lys(biotin)-OH (190 mg, 1.5 eq., 2 h), Fmoc-6-aminohexanoic acid (Fluka, 226 mg, 3 eq., 1 h), Fmoc-Tyr(tBu)-OH (460 mg, 5 eq., 1 h), Fmoc-Leu-OH (353 mg, 5 eq., 1 h) and ethyl-(2S, 3S)-oxirane-2,3-dicar-

boxylate (68 mg, 2 eq., 1 h) were coupled successively at indicated molar excess. The peptide was cleaved off the resin by addition of cleavage buffer (95 % TFA, 2.5% H_2O, 2.5 % triisopropylsilane) and precipitated as a white solid by ether at 0°C. The crude peptide was analysed by mass-spectrometry and used in subsequent experiments without further purification.

Cultivation of **Tetrahymena**

The *T. thermophila* strain CU438 was cultivated in a Bioengineering Kleinlaborfermenter on modified medium as described previously. Cell free supernatants were concentrated ten fold.

Protease activity assay

Protease activity was determined on microtiter plates by the substrate N-benzoyl-DL-arginine p-Nitroanilide (BAPNA): 50 μl sample were mixed with 200 μl buffer TED (200 mM Tris, 2 mM DTT, pH 7.5) and 10 μl BAPNA solution (20 mg/ml). A kinetic curve of the optical density at 410 nm was tracked on a microtiter plate reader for one hour. Papain (3.3 U/ml) served as reference and activity was calculated by linear regression of the recorded slope.

Purification of extracellular proteases

Cell free supernatants of *Tetrahymena* fermentation media were adjusted to 50 mM Tris pH 7.4 and 5 mM DTT and incubated with a final concentration of 0.2 mM DCG-04 for 2 h at room temperature. The samples were dialyzed against buffer B (50 mM Tris, 150 mM NaCl, pH 7.4). SDS was added to a final concentration of 0.5 % and samples were boiled for 10 min. Subsequently the samples were diluted with buffer B until the SDS concentration was as low as 0.2 % followed by shaking with pre-equilibrated streptavidin beads (Molecular Probes) at room temperature for one hour. The beads were thoroughly washed with buffer B, boiled in Laemmli-sample buffer and supernatants were subjected to SDS-PAGE. Protein bands were stained with Coomassie or silver.

Sequence determination of proteases

Protein bands of interest were cut out of the gel and tryptic in gel digestion was performed according to standard protocols. Samples were analysed on a set-up consisting of a
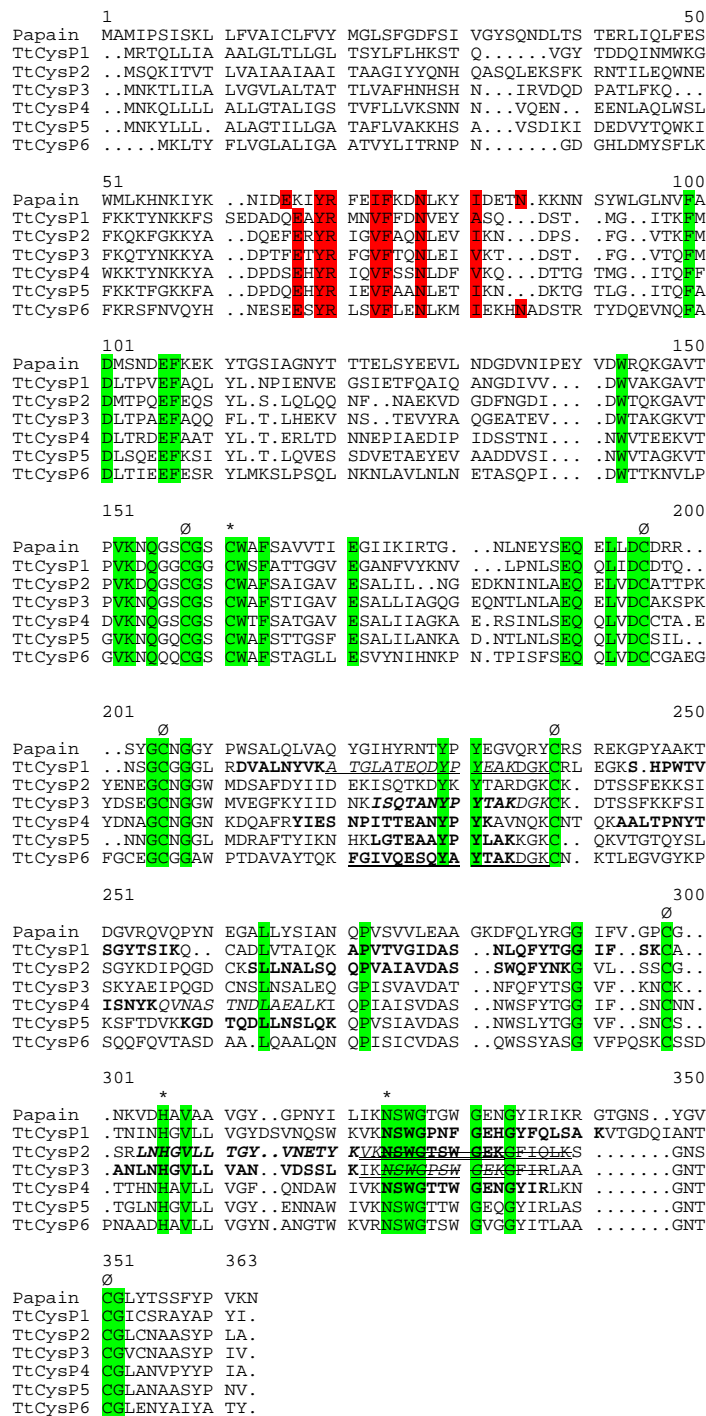
```
         1                                                            50
Papain   MAMIPSISKL LFVAICLFVY MGLSFGDFSI VGYSQNDLTS TERLIQLFES
TtCysP1  ..MRTQLLIA AALGLTLLGL TSYLFLHKST Q......VGY TDDQINMWKG
TtCysP2  ..MSQKITVT LVAIAAIAAI TAAGIYYQNH QASQLEKSFK RNTILEQWNE
TtCysP3  ..MNKTLILA LVGVLALTAT TLVAFHNHSH N...IRVDQD PATLFKQ...
TtCysP4  ..MNKQLLLL ALLGTALIGS TVFLLVKSNN N...VQEN.. EENLAQLWSL
TtCysP5  ..MNKYLLL. ALAGTILLGA TAFLVAKKHS A...VSDIKI DEDVYTQWKI
TtCysP6  .....MKLTY FLVGLALIGA ATVYLITRNP N.......GD GHLDMYSFLK

         51                                                          100
Papain   WMLKHNKIYK ..NIDEKIYR FEIFKDNLKY IDETN.KKNN SYWLGLNVFA
TtCysP1  FKKTYNKKFS SEDADQFAYR MNVFFDNVEY ASQ...DST. .MG..ITKFM
TtCysP2  FKQKFGKKYA ..DQEFFRYR IGVFAQNLEV IKN...DPS. .FG..VTKFM
TtCysP3  FKQTYNKKYA ..DPTFETYR FGVFTQNLEI VKT...DST. .FG..VTQFM
TtCysP4  WKKTYNKKYA ..DPDSEHYR IQVFSSNLDF VKQ...DTTG TMG..ITQFF
TtCysP5  FKKTFGKKFA ..DPDQEHYR IEVFAANLET IKN...DKTG TLG..ITQFA
TtCysP6  FKRSFNVQYH ..NESEFSYR LSVFLENLKM IEKHNADSTR TYDQEVNQFA

         101                                                         150
Papain   DMSNDEFKEK YTGSIAGNYT TTELSYEEVL NDGDVNIPEY VDWRQKGAVT
TtCysP1  DLTPVEFAQL YL.NPIENVE GSIETFQAIQ ANGDIVV... .DWVAKGAVT
TtCysP2  DMTPQEFEQS YL.S.LQLQQ NF..NAEKVD GDFNGDI... .DWTQKGAVT
TtCysP3  DLTPAEFAQQ FL.T.LHEKV NS..TEVYRA QGEATEV... .DWTAKGKVT
TtCysP4  DLTRDEFAAT YL.T.ERLTD NNEPIAEDIP IDSSTNI... .NWVTEEKVT
TtCysP5  DLSQEEFKSI YL.T.LQVES SDVETAEYEV AADDVSI... .NWVTAGKVT
TtCysP6  DLTIEEFRESR YLMKSLPSQL NKNLAVLNLN ETASQPI... .DWTTKNVLP

         151                                                         200
                   ∅      *                                      ∅
Papain   PVKNQGSCGS CWAFSAVVTI EGIIKIRTG. ..NLNEYSEQ ELLDCDRR..
TtCysP1  PVKDQGGCGG CWSFATTGGV EGANFVYKNV ...LPNLSEQ QLIDCDTQ..
TtCysP2  PVKDQGSCGS CWAFSAIGAV ESALIL..NG EDKNINLAEQ ELVDCATTPK
TtCysP3  PVKNQGSCGS CWAFSTIGAV ESALLIAGQG EQNTLNLAEQ ELVDCAKSPK
TtCysP4  DVKNQGSCGS CWTFSATGAV ESALIIAGKA E.RSINLSEQ QLVDCCTA.E
TtCysP5  GVKNQGQCGS CWAFSTTGSF ESALILANKA D.NTLNLSEQ QLVDCSIL..
TtCysP6  GVKNQQCGS CWAFSTAGLL ESVYNIHNKP N.TPISFSEQ QLVDCCGAEG

         201                                                         250
              ∅                                      ∅
Papain   ..SYGCNGGY PWSALQLVAQ YGIHYRNTVP YEGVQRYCRS REKGPYAAKT
TtCysP1  ..NSGCGGGL RDVALNYVKA TGLATEQDYP YEAKDGKCRL EGKS.HPWTV
TtCysP2  YENEGCNGGW MDSAFDYIID EKISQTKDYK YTARDGKCK. DTSSFEKKSI
TtCysP3  YDSEGCNGGW MVEGFKYIID NKISQTANYP YTAKDGKCK. DTSSFKKFSI
TtCysP4  YDNAGCNGGN KDQAFRYIES NPITTEANYP YKAVNQKCNT QKAALTPNYT
TtCysP5  ..NNGCNGGL MDRAFTYIKN HKLGTEAAYP YLAKKGKC.. QKVTGTQYSL
TtCysP6  FGCEGCGGAW PTDAVAYTQK FGIVQESQYA YTAKDGKCN. KTLEGVGYKP

         251                                                         300
                                                             ∅
Papain   DGVRQVQPYN EGALLYSIAN QPVSVVLEAA GKDFQLYRGG IFV.GPCG..
TtCysP1  SGYTSIKQ.. CADLVTAIQK APVTVGIDAS ..NLQFYTGG IF..SKCA..
TtCysP2  SGYKDIPQGD CKSLLNALSQ QPVAIAVDAS ..SWQFYNKG VL..SSCG..
TtCysP3  SKYAEIPQGD CNSLNSALEQ GPISVAVDAT ..NFQFYTSG VF..KNCK..
TtCysP4  ISNYKQVNAS TNDLAEALKI QPIAISVDAS ..NWSFYTGG IF..SNCNN.
TtCysP5  KSFTDVKKGD TQDLLNSLQK QPVSIAVDAS ..NWSLYTGG VF..SNCS..
TtCysP6  SQQFQVTASD AA.LQAALQN QPISICVDAS ..QWSSYASG VFPQSKCSSD

         301                                                         350
                  *                              *
Papain   .NKVDHAVAA VGY..GPNYI LIKNSWGTGW GENGYIRIKR GTGNS..YGV
TtCysP1  .TNINHGVLL VGYDSVNQSW KVKNSWGPNF GEHGYFQLSA KVTGDQIANT
TtCysP2  .SRLNHGVLL TGY..VNETY KVKNSWGTSW GEKGFIQLKS .......GNS
TtCysP3  .ANLNHGVLL VAN..VDSSL KIKNSWGPSW GEKSFIRLAA .......GNT
TtCysP4  .TTHNHAVLL VGF..QNDAW IVKNSWGTTW GENGYIRLKN .......GNT
TtCysP5  .TGLNHGVLL VGY..ENNAW IVKNSWGTTW GEQGYIRLAS .......GNT
TtCysP6  PNAADHAVLL VGYN.ANGTW KVRNSWGTSW GVGGYITLAA .......GNT

         351        363
            ∅
Papain   CGLYTSSFYP VKN
TtCysP1  CGICSRAYAP YI.
TtCysP2  CGLCNAASYP LA.
TtCysP3  CGVCNAASYP IV.
TtCysP4  CGLANVPYYP IA.
TtCysP5  CGLANAASYP NV.
TtCysP6  CGLENYAIYA TY.
```

Figure 6
Sequence alignment of *T. thermophila* extracellular proteases. The identified cysteine proteases in *T. thermophila* supernatants are compared to papain. The ERFNIN motif is highlighted in red, amino acids that are conserved in all enzymes are green. Amino acids of the catalytic center are marked by *, cysteines that form disulfide bridges by ∅. All peptides found by mass spectrometry in the different proteases are highlighted by either bold, italic, underlined, or doubly underlined letters. Obviously no N-terminal peptides were detected, as only secreted, mature enzymes with cleaved off prepro-peptides (~aa 1-140) were analysed. Peptides around aa 160 are not expected either as the catalytic cysteine that covalently bound to DCG-04 resides at this position.

Table 2: Putative cysteine proteases in the *T. thermophila* genome

Name	SignalP				ERFNIN motif					
	propability [%]	E	$X_{2/3}$	R	(I/V)	F	N	(I/V)	$X_{1/2/3}$	N
136.m00075	79	E	3	R	V	F	N	-	-	-
295.m00026	73	E	2	R	V	F	N	I	2	N
295.m00023	98	E	2	R	I	F	N	V	2	N
1.m00874	97	E	3	R	V	F	N	I	-	-
103.m00130	93	E	2	R	V	F	N	I	-	-
5.m00542	98	E	2	R	V	F	N	V	I	N
24.m00271	99	E	2	R	I	F	N	I	-	-
5.m00535	98	E	2	R	V	F	N	V	-	-
147.m00097	98	E	2	R	V	F	N	I	-	-
120.m00105	99	E	2	R	V	F	N	V	I	N
150.m00098	96	E	3	R	I	F	N	I	3	N
89.m00111	99	E	3	R	V	F	N	I	3	N
304.m00023	100	-	-	R	I	F	N	I	3	N
13.m00464	93	-	-	R	I	F	N	V	3	N
61.m00268	96	-	-	R	I	F	N	I	-	-
9.m00484	100	E	3	R	V	F	N	V	3	N
9.m00485	100	E	3	R	V	F	N	I	3	N
9.m00486	100	-	-	R	V	F	N	I	3	N
98.m00123	98	E	3	R	V	F	N	V	3	N
98.m00125	98	E	3	R	V	F	N	V	3	N
98.m00126	98	E	3	R	V	F	N	V	3	N
98.m00131	100	E	3	R	V	F	N	V	3	N
98.m00128	98	E	3	R	V	F	N	V	3	N
156.m00128	100	E	3	R	V	F	N	I	3	N
156.m00129	100	E	3	R	V	F	N	I	3	N
343.m00015	100	E	3	R	V	F	N	I	3	N
343.m00014	100	E	3	R	V	F	N	I	3	N
156.m00109	100	E	3	R	V	F	N	I	3	N
5.m00554	100	E	3	R	I	F	N	V	3	N
49.m00253	100	E	3	R	V	F	N	I	-	-
218.m00064	100	E	3	R	L	F	N	I	3	N

Dionex Ultimate HPLC equipped in combination with a Famos autosampler, a Switchos switcher and a ThermoElectron LTQ mass spectrometer. Peptides were identified using Sequest software fed with *T. thermophila* preliminary protein sequence data, that was obtained from The Institute for Genomic Research website. [30] Sequence alignments were performed with Multalin version 5.4.1 [31] (Symbol comparison table: blosum62, Gap weight: 12, Gap length weight: 2, Consensus levels: high = 90% low = 50%).

Authors' contributions
LH synthesized DCG-04, participated in project conception and drafted the manuscript. ME carried out the purification, SDS gels and enzyme assays and evaluated MS and sequencing data. IA developed protease assays and supervised fermentation processes. AT helped to draft the manuscript. MWWH conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

References
1. Greider CW, Blackburn EH: **Identification of a specific telomere terminal transferase activity in Tetrahymena extracts.** *Cell* 1985, **43**:405-413.
2. Blackburn EH, Gall JG: **A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena.** *J Mol Biol* 1978, **120**:33-53.
3. Cech TR, Zaug AJ, Grabowski PJ: **In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence.** *Cell* 1981, **27**:487-496.
4. Pan WC, Blackburn EH: **Single extrachromosomal ribosomal RNA gene copies are synthesized during amplification of the rDNA in Tetrahymena.** *Cell* 1981, **23**:459-466.
5. Gaertig J, Gorovsky MA: **Efficient mass transformation of Tetrahymena thermophila by electroporation of conjugants.** *Proc Natl Acad Sci U S A* 1992, **89**:9196-9200.
6. Cassidy-Hanley D, Bowen J, Lee JH, Cole E, VerPlank LA, Gaertig J, Gorovsky MA, Bruns PJ: **Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment.** *Genetics* 1997, **146**:135-147.

7. Larson DD, Blackburn EH, Yaeger PC, Orias E: **Control of rDNA replication in Tetrahymena involves a cis-acting upstream repeat of a promoter element.** *Cell* 1986, **47:**229-240.
8. Hai B, Gaertig J, Gorovsky MA: **Knockout heterokaryons enable facile mutagenic analysis of essential genes in Tetrahymena.** *Methods Cell Biol* 2000, **62:**513-531.
9. Gaertig J, Kapler G: **Transient and stable DNA transformation of Tetrahymena thermophila by electroporation.** *Methods Cell Biol* 2000, **62:**485-500.
10. Clark TG, Gao Y, Gaertig J, Wang X, Cheng G: **The I-antigens of Ichthyophthirius multifiliis are GPI-anchored proteins.** *J Eukaryot Microbiol* 2001, **48:**332-337.
11. Peterson DS, Gao Y, Asokan K, Gaertig J: **The circumsporozoite protein of Plasmodium falciparum is expressed and localized to the cell surface in the free-living ciliate Tetrahymena thermophila.** *Mol Biochem Parasitol* 2002, **122:**119-126.
12. Wheatley DN, Rasmussen L, Tiedtke A: **Tetrahymena: a model for growth, cell cycle and nutritional studies, with biotechnological potential.** *Bioessays* 1994, **16:**367-372.
13. Kiy T, Tiedtke A: **Continuous high-cell-density fermentation of the ciliated protozoon Tetrahymena in a perfused bioreactor.** *Appl Microbiol Biotechnol* 1992, **38:**141-146.
14. Kiy T, Tiedtke A: **Effects of immobilization on growth, morphology, and DNA content of the ciliated protozoon Tetrahymena thermophila.** *FEMS Microbiol Lett* 1993, **106:**117-122.
15. Banno Y, Sasaki N, Nozawa Y: **Secretion heterogeneity of lysosomal enzymes in Tetrahymena pyriformis.** *Exp Cell Res* 1987, **170:**259-268.
16. Alam S, Banno Y, Nozawa Y: **Purification and characterization of phospholipase C preferentially hydrolysing phosphatidylcholine in Tetrahymena membranes.** *J Eukaryot Microbiol* 1993, **40:**775-781.
17. Kovacs P, Csaba G, Nakashima S, Nozawa Y: **Phospholipase D activity in the Tetrahymena pyriformis GL.** *Cell Biochem Funct* 1997, **15:**53-60.
18. Guberman A, Hartmann M, Tiedtke A, Florin-Christensen J, Florin-Christensen M: **A method for the preparation of Tetrahymena thermophila phospholipase A1 suitable for large-scale production.** *J Appl Microbiol* 1999, **86:**226-230.
19. Banno Y, Nozawa Y: **Purification and characterization of lysosomal alpha-glucosidase secreted by eukaryote Tetrahymena.** *J Biochem (Tokyo)* 1985, **97:**409-418.
20. Rasmussen L, Florin-Christensen M, Florin-Christensen J, Kiy T, Tiedtke A: **Differential increase in activity of acid phosphatase induced by phosphate starvation in Tetrahymena.** *Exp Cell Res* 1992, **201:**522-525.
21. Straus JW, Migaki G, Finch MT: **An assessment of proteolytic enzymes in Tetrahymena thermophila.** *J Protozool* 1992, **39:**655-662.
22. Suzuki KM, Hosoya N, Takahashi T, Kosaka T, Hosoya H: **Release of a newly-identified cysteine protease, tetrain, from Tetrahymena into culture medium during the cell growth.** *J Biochem (Tokyo)* 1997, **121:**642-647.
23. Karrer KM, Peiffer SL, DiTomas ME: **Two distinct gene subfamilies within the family of cysteine protease genes.** *Proc Natl Acad Sci U S A* 1993, **90:**3063-3067.
24. van der Hoorn RA, Leeuwenburgh MA, Bogyo M, Joosten MH, Peck SC: **Activity profiling of papain-like cysteine proteases in plants.** *Plant Physiol* 2004, **135:**1170-1178.
25. Greenbaum DC, Baruch A, Grainger M, Bozdech Z, Medzihradszky KF, Engel J, DeRisi J, Holder AA, Bogyo M: **A role for the protease falcipain 1 in host cell invasion by the human malaria parasite.** *Science* 2002, **298:**2002-2006.
26. Kocks C, Maehr R, Overkleeft HS, Wang EW, Iyer LK, Lennon-Dumenil AM, Ploegh HL, Kessler BM: **Functional Proteomics of the Active Cysteine Protease Content in Drosophila S2 Cells.** *Mol Cell Proteomics* 2003, **2:**1188-1197.
27. Orias E: **Toward sequencing the Tetrahymena genome: exploiting the gift of nuclear dimorphism.** *J Eukaryot Microbiol* 2000, **47:**328-333.
28. Bendtsen JD, Nielsen H, von HG, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340:**783-795.
29. Stargell LA, Karrer KM, Gorovsky MA: **Transcriptional regulation of gene expression in Tetrahymena thermophila.** *Nucleic Acids Res* 1990, **18:**6637-6639.
30. **The Institute for Genomic Research.** [http://www.tigr.org].
31. Corpet F: **Multiple sequence alignment with hierarchical clustering.** *Nucleic Acids Res* 1988, **16:**10881-10890.