Cloning, expression, purification, and characterisation of the HEAT-repeat domain of TOR from the thermophilic eukaryote *Chaetomium thermophilum*

Graham C. Robinson^{a*}, Yogesh Vegunta^a, Caroline Gabus^a, Christl Gaubitz^a, and Stéphane Thore^{ac†}

^aDepartment of Molecular Biology, University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva, Switzerland

^bEuropean Molecular Biology Laboratory, Grenoble Outstation, 6 Rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France

^cINSERM U-1212, CNRS UMR-5320, Université de Bordeaux, 146 rue Léo Saignat, 33000 Bordeaux, France

^{*}Current address: Department of Botany and Plant Biology, University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva, Switzerland

[†]Corresponding Author: stephane.thore@inserm.fr

Abbreviations

TOR(C), Target of rapamycin (complex); MOI, multiplicity of infection; PBS, phosphate buffered saline; HEAT-repeat, Huntingtin, Elongation Factor 3, Protein Phosphatase 2A, and TOR1; FAT, FRAP-ATM-TRRAP; FAT-C, FRAP-ATM-TRRAP C-terminal; FRB, FKBP12-rapamycin binding domain; PI3K, phosphatidyl inositol 3-kinase; LBE, mLST8 Binding Element; CDS, coding sequence; EDTA, Ethylenediaminetetraacetic acid; EGTA, ethylene glycol-bis(β-aminoethyl ether)tetraacetic acid; DTT, Dithiothreitol; PMSF, phenylmethanesulfonyl fluoride; SDS, sodium dodecylsulphate; TEVp, Tobacco Etch Mosaic Virus protease; SEC, size exclusion chromatography; MALDI-TOF, matrix-assisted laser desorption ionization-time of flight; MS, mass spectrometry

Keywords (*maximum of 6*)

TOR; *Chaetomium thermophilum*; mass spectrometry; target of rapamycin signalling; purification; MultiBac expression system

Graphical Abstract

Abstract

The Target of Rapamycin Complex is a central controller of cell growth and differentiation in eukaryotes. Its global architecture has been described by cryoelectron microscopy, and regions of its central TOR protein have been described by X-ray crystallography. However, the N-terminal region of this protein, which consists of a series of HEAT repeats, remains uncharacterised at high resolution, most likely due to the absence of a suitable purification procedure. Here, we present a robust method for the preparation of the HEAT-repeat domain, utilizing the thermophilic fungus *Chaetomium thermophilum* as a source organism. We describe construct design and stable expression in insect cells. An efficient two-step purification procedure is presented, and the purified product is characterised by SEC and MALDI-TOF MS. The methods described pave the way for a complete high-resolution characterisation of this elusive region of the TOR protein.

Introduction

The Target Of Rapamycin Complex (TORC) is the nexus of its eponymous signalling pathway, which controls cellular growth and differentiation in eukaryotes. It was discovered with a yeast genetic screen, which identified mutations conferring resistance to the antifungal agent rapamycin [1], and became the founding member of the phosphatidylinositol-kinase-related kinase (PIKK) family[2]. There are two different TORCs: TORC1 and TORC2, which regulate temporal and spatial aspects of cell growth, respectively [3, 4]. The complexes differ in their subunit compositions, but both are constructed around a large, multi-domain TOR protein, containing the PI3K-like catalytic domain and associated regulatory regions [5]. The two complexes are defined by their associated subunits, which have been reviewed in detail elsewhere [6-9].

The TORCs are a key focus for clinical study due to their roles as central controllers of cell growth and differentiation, as well as their function in the immunosuppressive and anti-proliferative effects of the antifungal agent rapamycin. Major strides have been taken towards the structural characterisation of the complexes. Electron microscopy structures are available for TORC1 and its subcomplexes [10-13], and TORC2 [14]. A 3.2 Å X-ray crystal structure is available for a subcomplex containing the kinase domain and FKBP12-binding domain of the human TOR protein [15], and structures of other TORC proteins and domains have been solved as well [13, 16-19].

However, at present there is no high-resolution structural information for the N-terminal region of the TOR protein. This region consists of a substantial HEAT-repeat domain ([5, 20-22], which forms an extended α -solenoid scaffold [11, 13], and is proposed to confer signalling functionality to TORCs through its interaction with upstream and downstream effector proteins [6, 23]. A complete structural characterisation of this region will be essential for a comprehensive understanding of the function and modulation of TOR signalling.

Although the quaternary structure of the HEAT-repeat domain of TOR has been described by electron microscopy [11, 13], the absence of atomic-resolution information for this region is indicative of its instability in the absence of binding partners. Here, we explore the use of a thermophilic fungus, *Chaetomium thermophilum*, as a source organism for structural study of the TOR protein.

First described in 1950 [24], *C. thermophilum* grows at elevated temperatures found in compost heaps and animal dung [24, 25], and is becoming a popular source organism for proteins suitable for structural studies [13, 26-28]. It can grow at temperatures up to 60°C [26], and can be cultured under laboratory conditions with media containing widely available components, and within timeframes amenable for research. *C. thermophilum* can be stored and propagated from spores or vacuum-dried cultures. Its genome has been sequenced, and partially annotated [26], facilitating protein engineering. Importantly, its proteins exhibit sequence homology to human proteins, making this organism of key interest as a model for structural study. The TOR protein from *C. thermophilum* is comparable to the other model TOR signalling systems (Table 1).

Species	Protein Name	Accession Number	Coverage	Sequence Identity	Total Score
<i>C. thermophilum</i>	TOR	G0S9M0	95	45	2028
<i>S. cerevisiae</i>	TOR1	P35169	95	42	1846
<i>S. cerevisiae</i>	TOR2	P32600	96	43	2004
<i>S. pombe</i>	TOR1	O14356	94	40	2063
<i>S. pombe</i>	TOR2	Q9Y7K2	96	46	2100
<i>D. discoideum</i>	TOR	Q86C65	93	48	2468

Table 1. Sequence homology of TOR proteins from model organisms relative to human mTOR (accession number P42345). Pairwise alignment was performed with Clustal Omega [29, 30].

Here, we have expressed, purified, and characterised the HEAT-repeat domain of the TOR protein (referred to here as TOR^{AC}) from *C. thermophilum*, the region of this keystone signalling protein that has so far eluded structural characterisation by X-ray crystallography.

Materials and Methods

Apparatus and Reagents

Reagents were purchased from Sigma-Aldrich (Wassergrasse 7, 9000 St. Gallen, Switzerland) and VWR (Rue Neuve 7, 1260 Nyon, Switzerland), unless described otherwise. Oligonucleotides were prepared by MicroSynth (MicroSynth AG, Balgach, Switzerland), and DNA sequencing was performed by the same company. Proteins were concentrated with spin concentrators with a 100,000 Da molecular weight cut-off (catalogue numbers: 810024/910024; Merck Millipore, Schaffhausen, Switzerland), and protein concentrations were estimated with the Bradford assay [31] (cat# 500-0006; Bio-Rad Laboratories AG, Cressier, Switzerland). Affinity chromatography media were purchased from IBA (Goettingen, Germany), and gel filtration chromatography was performed with a Superdex200 (10/300) GL (cat# 17-5175-01; GE Healthcare, Uppsala, Sweden).

Growth of Chaetomium thermophilum

Vacuum-dried cultures of *C. thermophilum* var. *thermophilum* (DSM 1495) were obtained from DSMZ Braunschweig, Germany. Minimal agar (DSMZ) consisting of 0.2% (w:v) NaNO₃, 0.05% (w:v) MgSO₄, 0.05% (w:v) KCl, 0.001% (w:v) Fe₂(SO₄)₃, 0.014% (w:v) KH₂PO₄, 0.12% (w:v) K₂HPO₄, 0.002% (w:v) yeast extract, and 1.5% (w:v) agar, pH 7.2, was inoculated with the dried cultures, and *C. thermophilum* was cultured at 50°C until sporulation. Spores were recovered from the plates in 20% glycerol, and stored at -70°C. For mycelial growth, spores were distributed on agar containing potato dextrose and yeast extract, consisting of 0.4% (w:v) potato infusion, 2% dextrose, and 0.002% yeast extract, and incubated at 52°C. After 4 days the surface of the agar was covered in *C. thermophilum* mycelium, which was recovered with a P1000 pipette tip, and transferred to a 2 ml Eppendorf tube for RNA extraction.

RNA extraction and cDNA preparation

Mycelium (100 mg) was suspended in 400 µl 50 mM CH₃CO₂Na, pH 5.3, 10 mM EDTA, 1% SDS, and 1 mg/ml proteinase K, and homogenised with a microspatula. The suspension was transferred to a 2 ml tube containing 200 µl glass beads and 400 µl acid-buffered phenol, and disrupted in an orbital shaker (2,200 rpm; 10 mins). Sample was incubated at 65°C for 5 mins, cooled on ice for 10 mins, and the supernatant was recovered by centrifugation (13,000 rpm; 2 mins). An equal volume of phenol:chloroform:isoamylalcohol (25:24:1, made with acid-buffered phenol) was added to the supernatant. The mixture was vortexed, and the soluble fraction (approximately 300 µl) was recovered by centrifugation (13,000 rpm; 2 mins). Then 30 µl of 3 M CH₃CO₂Na, pH 5.2, and 750 µl ethanol were added. The mixture was incubated at -70°C for 1 hr, and precipitated material is recovered by centrifugation (13,000 rpm; 2 mins). The recovered pellet was washed with 80% ethanol, and resuspended in 30 µl of water. Background genomic DNA was removed with an RNeasy Minikit (cat# 74104, Qiagen AG, Garstligweg 8, 8634 Hombrechtikon, Switzerland), and the quality of recovered RNA was assessed by analysing ribosomal RNA by agarose gel electrophoresis with a 1 kb ladder (cat# N3232L, New England Biolabs, 240 County Road, Ipswich, MS, US). First-strand cDNA was synthesised with a PrimeScript First Strand cDNA Synthesis Kit (cat# 6110A, Takara Clontech, 2 Avenue du President Kennedy, 78100 Saint-Germain-en-Laye, France) using oligo dT primers.

Protein engineering and constructs

Secondary and tertiary structure predictions of the HEAT-repeat domain of TOR were prepared with Phyre² [32], and oligonucleotide primers were designed for cloning the coding region corresponding to residue 1 to 1095. Despite an excellent cDNA library, we failed to generate the cDNA corresponding to the coding sequence encompassing residues 1-1095 of TOR. Using genomic DNA from *C. thermophilum*, we PCR amplified the coding sequence of TOR from nucleotides 772 to 2580 (primer Tor-772-For and Tor-2580-2654-Rev) and 2654 to 4021 (primers Tor-2580-2654-For and Tor-3896-4000-Rev). This first step of PCR allowed the removal of a short intron located between the nucleotide 2580 and 2654 of the genomic sequence. In a second round of PCR, both fragment of sequences were fused together (primers Tor-2580-2654-Rev and Tor-2580-2654-For have over lapping sequences) and the fused sequence cloned into a pCR'-Blunt-TOPO vector (Invitrogen). Because multiple introns reside in the 5'-region of the TOR gene, we ordered a synthetic sequence (from Entelechon) coding for the first 259 residues of the TOR protein (nucleotide 1-777). The synthetic sequence was amplified by PCR using the primers Tor-syn-SpeI-1-For and Tor-syn-BsrGI-800-Rev. The PCR amplified synthetic sequence and the pCR'-Blunt-TOPO vector containing the genomic sequence corresponding to nucleotide 772-4021 were subsequently digested with

the restriction enzyme SpeI and BsrGI and ligated together. The fragment coding for the first 1095 amino acids of TOR was finally amplified by PCR using the primers SLIC-TOR1-For and SLIC-TOR1095-Rev and introduced into *MultiBac^{Turbo}* with the ACEMBL expression system using the primer SLIC-pACEB-For and SLIC-pACEB-Rev [33-37]. Table 2 contains the sequence of the primers used for the various cloning steps.

Primer name	Sequence
Tor-772-For	AGGCGGCATGTACATGCAGAGTCACTAC
Tor-2580-2654-Rev	CGCTCTCTCTTCAACCTGCTGGTATTTATAGGGATC
Tor-2580-2654-For	GATCCCTATAAATACCAGCAGGTTGAAGAGAGAGCG
Tor-3896-4000-Rev	CGATGTTGGTGATAAGGTCCTCCTGATGCTGTTTCGTAC
Tor-syn-SpeI-1-For	acgtactagtacATGTCATCTGTAATCCG
Tor-syn-BsrGI-800-Rev	GTAGTGACTCTGCATGTACATGCCACCCT
SLIC-TOR1-For	gagaacctgtactccagATGTCATCTGTAATCCGCGATAACC
SLIC-TOR1095-Rev	gattcgaaagtcgattcgatcctaGCCCTTTTGGAGCTTC
SLIC-pACEB-For	taggatcgaatcggactttcgaatctagagcctgc
SLIC-pACEB-Rev	ctggaagtacagggttctcgcc

Table 2. List of primers and associated sequences used for the construction of the Tor^{ΔC}.

Expression of TOR^{ΔC}

Proteins were expressed in Sf9 cells according to methods described previously [36]. Briefly, confluent Sf9 cells at a density of 8×10^5 /ml were diluted 1:2 (v:v) with SF-9 BaculoExpress media (cat# 9-00F38-I, Bioconcept, Paradiesrain 14, 4123 Allschwil, Switzerland). *MultiBac^{Turbo}* containing TOR^{ΔC} was recovered from DH10 *MultiBac^{Turbo}* cells by alkaline lysis, dissolved in 100 μ l SF-9 BaculoExpress media containing 5 μ l FuGene transfection reagent (cat# E2311, Promega, Wallisellenstrasse 55, 8600 Dübendorf, Switzerland) and incubated at 25°C (25 mins). This transfection mixture was added to the cell culture and incubated at 27°C. Infection was monitored by fluorescence of a marker YFP incorporated into *MultiBac^{Turbo}*. Transfection was complete after 48-72 hr, and V₀ virus was recovered by filtration. Virus was amplified by infecting 200 ml insect cells (8×10^5 /ml) with 250 μ l V₀ (MOI<1). Once transfection was complete cells were removed by centrifugation (1,500 x g; 20 mins), and V₁ was stored at 4°C. For large scale expression, 1 L insect cell culture (3×10^6 /ml) was infected with 30 ml V₁. Once transfection was complete (usually 48-72 hr) cells were collected by centrifugation (1,500 x g; 20 mins), and resuspended (10 ml/Ltr culture) in PBS containing a Complete Protease Inhibitor Cocktail (cat# 11697498001, Roche Diagnostics, Basel, Switzerland). Cell suspensions were dripped into liquid nitrogen, forming cell beads, which were stored at -80°C.

Purification of TOR^{ΔC}

Expression cells (10 g) were thawed at 4°C in 35 ml resuspension buffer composed of 100 mM Tris·HCl, 10% sucrose, 50 mM NH₄Cl, 10 mM DTT, 1 mM EDTA, 1 mM EGTA, 5 mM PMSF, pH 8.0. A Complete Protease Inhibitor Cocktail tablet and 1 ml BioLock Biotin Blocking Solution (cat# 2-0205-050, IBA, Goettingen, Germany) were added, and the suspension was sonicated on ice with a microtip (8 x 30s+30s pause at an amplitude of 20%; Branson Digital Sonifier 450, Branson Ultrasonics Corp., Danbury, CT, USA). The following steps were performed at 4°C. Unbroken cells and insoluble debris were removed by centrifugation (47,000 x g; 20 mins), and the clarified lysate was applied to a 5 ml StepTactin column (cat# 2-1234-001, IBA, Goettingen, Germany) at 1 ml/min with a

peristaltic pump. The column was washed with resuspension buffer until the eluate produced no visible colour change when protein presence was assessed with the Bradford Reagent test [31]. Immobilised TOR^{ΔC} was released from the column with 1 mg TEV protease in 5 ml of resuspension buffer, and incubated for 16 hr. TOR^{ΔC} was recovered from the column in the same buffer, and elution was monitored with Bradford Reagent [31]. Protein was concentrated to 200 μl, and applied to a Superdex200 (10/300) gel filtration column (0.4 ml/min) equilibrated with 20 mM Tris·HCl, 10% sucrose, 50 mM NH₄Cl, 10 mM DTT, pH 8.0. Elution was monitored by absorbance at 280 nm. The purity of TOR^{ΔC} was analysed by SDS-PAGE (10%; [38], and visualised with Coomassie Brilliant Blue. Subunit molecular masses were estimated by comparison to electrophoresis standards (Page Ruler Plus, cat# 26619; ThermoFisher Scientific, LuBioScience GmbH, Topterstrasse 5, 6000 Lucerne, Switzerland). Protein yield was estimated by Bradford Assay [31], with a standard curve prepared with bovine serum albumin. Purity was estimated by densitometry with ImageJ [39].

Molecular Mass Estimation

An estimation of molecular mass was obtained from the Superdex200 (10/300) column. The elution profile of the column was calibrated with native protein standards (thyroglobulin, ferritin, ovalbumin, and monomeric bovine serum albumin), and the void volume of the column was determined with blue dextran.

Peptide mass fingerprinting

Bands containing purified protein were cut from an SDS-PAGE gel, and washed thoroughly with water to remove destaining solution. Peptides were prepared by in-gel tryptic digestion incorporating in-gel reduction and alkylation [40]. Peptides were extracted with 25 mM ammonium bicarbonate, then 5% formic acid, and the resulting solutions were concentrated in a speedvac at room temperature. Drops (0.6 μl) of a 1:1 solution with 2,5-dihydroxybenzoic acid (40 mg/ml) were applied to a MTB 384 target plate (cat# 8280784/8074115, Bruker, Permoserstrasser 15, 04318 Leipzig, Germany). Mass spectra were obtained by MALDI-TOF with a Bruker Autoflex Speed calibrated with Peptide Calibration Standards II (cat# 8222570, Bruker). Monoisotopic mass peaks were analysed with MASCOT [41] searches of the NCBIprot database. Searches allowed for two missed trypsin cleavages, a peptide tolerance of +/-0.1 Da, carbamidomethyl modification (fixed modification) and methionine oxidation (variable modification).

Results and Discussion

Preparation of total cell RNA from Chaetomium thermophilum

C. thermophilum grew readily as mycelium on potato dextrose agar containing yeast extract (Figure 1.A), and was recovered with a pipette tip. Cell extract was obtained from the mycelium with acid-buffered SDS in an orbital shaker, and intact total-cell RNA was recovered by phenol:chloroform:isoamylalcohol extraction. Fidelity of the extraction was assessed by agarose gel electrophoresis (Figure 1.B).

Figure 1. Growth and RNA extraction from *Chaetomium thermophilum*. **A**, Mycelial growth of *Chaetomium thermophilum* on potato dextrose agar containing yeast extract. **B**, Gel electrophoresis profile of RNA extract. Band sizes for the ladder are given in kb. 26 S and 18 S rRNA bands are indicated.

Construction and expression of TOR^{ΔC}

In common with metazoans, *C. thermophilum* contains a single copy of the TOR protein. A 1095 residue region of the HEAT-repeat domain, referred to here as TOR^{ΔC}, was designed such that predicted secondary and tertiary structural elements were not disrupted (Figure 2). Having failed to produce numerous TOR constructs as well as other proteins known to associate with the TOR protein to form the TOR complex and given the size of TOR^{ΔC} construct (~123 kDa), we tried to express it in insect cells using the MultiBac technology. Thus, the CDS for TOR^{ΔC} was cloned into the pACEBac1 vector, introduced into the Baculovirus genome using MultiBac technology and expressed in Sf9 cells. Typical cell yield under expression conditions was approximately 20 g/L.

Figure 2. Domain architecture of TOR and TOR^{ΔC}. TOR is composed of HEAT, FAT, and kinase domains. The kinase domain is further divided into FRB, LBE, and FATC domains. TOR^{ΔC} is indicated below. Numbers indicate amino acid residues. Almost all of the HEAT-repeat region remains uncharacterised by X-ray crystallography (grey).

Purification of TOR^{ΔC}

Detergent-free cell lysis was achieved by mild sonication, and the clarified cell lysate was applied to a StrepTactin column. TOR^{ΔC} was recovered from the column by TEV protease cleavage of the affinity tag. Due to the highly selective nature of StrepTactin binding and TEV protease cleavage, no tangible improvement in purity was obtained by fractionation at this step. Consequently, the purification procedure was streamlined by application of a peristaltic pump.

Protein was concentrated, and applied to a Superdex200 (10/300) column. Monodisperse TOR^{ΔC} was separated from aggregated material and TEVp (Figure 3). It should be noted that under commonly-used purification conditions TOR^{ΔC} could not be purified in a non-aggregated state, as determined by gel filtration chromatography and confirmed with MALS (*data not shown*). Electron microscopy structures of the TOR complex reveal that the HEAT-repeat domain of the TOR protein provides a substantial interaction interface with other TORC proteins [11, 13, 14]. Consequently, it is not unexpected that TOR^{ΔC} lacks stability when prepared in isolation. Through compound screening we found that incorporation of 10% (w:v) sucrose and 50 mM NH₄Cl were sufficient for recovery of monodisperse TOR^{ΔC}.

A final yield of 0.8 mg of TOR^{ΔC} was obtained from 10 g expression cells (bead weight). The product is near-homogeneous with an approximate purity of 93%. Due to lack of enzymatic activity of TOR^{ΔC} purification efficiency could not be assessed by specific activity. Instead, the purity of TOR^{ΔC} was obtained by densitometry.

Figure 3. Purification of TOR^{ΔC}. **A**, Elution profile from S200 (10/300). Peak elution volume and fractions pooled for identification and quantification (—) are indicated. **B**, SDS-PAGE analysis of preparation. Fraction numbers are indicated above gel. Molecular masses of electrophoresis standards (S) are presented in kDa. **C**, Standard curve for size exclusion calibration standards. The elution position of TOR^{ΔC} (o) is indicated.

Characterisation of TOR^{ΔC}

The elution volume of TOR^{ΔC} from gel filtration indicates a molecular mass of 126,000 Da (Figure 3). This value corresponds closely to the molecular mass calculated from primary sequence (123,263 Da), and suggests that TOR^{ΔC} is recovered in a monomeric form under the conditions described. Its identity was confirmed by peptide mass fingerprinting (Figure 4. Characterisation of TOR^{ΔC}. **A**, MALDI-

TOF MS profile of TOR^{ΔC} tryptic digest. **B**, Sequence of TOR^{ΔC}. Sequence coverage of peptide hits returned from MASCOT (**Table**) are bold and underlined.

and Table). Phosphatidylinositol 3-kinase-like protein from *Chaetomium thermophilum* (accession: EGS20131.1) was returned as the only statistically significant hit ($p < 0.05$) from the NCBIprot database. Sequence coverage of 17% was obtained for TOR^{ΔC}.

B

1	11	21	31	41	51
MSSVIRDTVS	SALDQYIKDA	RSRSHGEDQR	KRAARQVREL	<u>VHIAKQEMSA</u>	<u>BQQRFFDTV</u>
61	71	81	91	101	111
<u>HQKTNALIQS</u>	TDTYDMGGV	YILDALVDFD	GIEPALKLSR	<u>POQYIGTILR</u>	GRDLNPMQPA
121	131	141	151	161	171
AVVLGKLCRP	GGSLISELVD	AEVQTALENM	QSDRVEERRY	<u>SAVLVIRELA</u>	PNCPTLMYPY
181	191	201	211	221	231
VGFFVDHIWI	GLRDPRLIR	ATSSETVSAC	FKIIRERDQD	LKQENMDRMF	AEAIGLKTN
241	251	261	271	281	291
SVEYIHASLL	VLKELLEQGG	MYMQSHYQEA	CDIVYRHKDA	RDPAIRKTVV	FLIPDLANYA
301	311	321	331	341	351
PSEFSATYLH	KFMVYLNNML	KKEKERNDAF	LAIGNIANSV	<u>KSSITPYLDS</u>	<u>ILVHVREGLS</u>
361	371	381	391	401	411
VQSRKRGSDV	PVFDCISRLA	VAVGQTLISKY	MEALLDPIFA	CELTPKLTQA	LVDMAFYIPP
421	431	441	451	461	471
VKGIIQERLL	NMLSEKVGGE	<u>PFRPLGAPYS</u>	<u>TPIPSIPPIP</u>	KDPKDSVHE	SGKAQVRLAL
481	491	501	511	521	531
NLGSFDFSG	MIINEFVRDV	AIKYVEDDDP	EVREAAALTC	CQLYIRDPIV	NQTSYNHALQV
541	551	561	571	581	591
VAEVVDKLLT	VGISDPEPKI	<u>RQTVLTALDE</u>	<u>RFDQHLAKAE</u>	<u>NIRTLFFALH</u>	<u>DEQFAVREAA</u>
601	611	621	631	641	651
VTIIGRLTRH	<u>NPATVIPHLR</u>	KTVIQMLTEL	EYTDVPSRKE	ESARLLSLLT	QHAQDLIKPY
661	671	681	691	701	711
VSSITEVLLP	KASDPTPSVA	ATILQAIGEL	CAVGAEMLA	YKDTLMPIII	DALQDQSAPI
721	731	741	751	761	771
KREAAHLTLG	QLASNAGYVI	KPYLEYFQLL	EILQSIIRGE	FQHGFLRQET	<u>IKLLGILGAL</u>
781	791	801	811	821	831
<u>DPTKYQQVEE</u>	<u>RAPQTQRPE</u>	ATQQTDSLSM	MSGITPSQED	YYPTVVINAL	LHILQDQSLV
841	851	861	871	881	891
QNHGNVVDIAI	MSIFITLGLR	CVQFLDRVVP	<u>AFIQVIRASS</u>	PSRLEFYFNH	LSRLVTIVRQ
901	911	921	931	941	951
HIRVYLPDII	ALLQEFWNIS	SSLQTTILNL	IESIARSLEG	EFKIIYLAGLL	FLMLGVLEKD
961	971	981	991	1001	1011
VSTERQPTER	IFHAPLVFGS	SAEEYMHLII	PILVRLFDSP	<u>AQPAFLRKSA</u>	IETIGKLSSM
1021	1031	1041	1051	1061	1071
VNLNDYASKI	INPLTRILAQ	<u>QEPSLRVAAL</u>	DTICAMMLQL	GRDYLHPERT	VDKIITMYGI
1081	1091				
QHANYDKAVE	KLRKG				

Figure 4. Characterisation of TOR^{AC}. **A**, MALDI-TOF MS profile of TOR^{AC} tryptic digest. **B**, Sequence of TOR^{AC}. Sequence coverage of peptide hits returned from MASCOT (**Table**) are bold and underlined.

Peak	Observed	Mr (Expected)	Mr (Calculated)	Mass Difference	Start	End	Miss	Peptide
1	920.5880	919.5807	919.5491	0.0316	160	167	0	R.YSAVLVLR.E
2	951.4970	950.4897	950.4457	0.0440	785	791	0	K.YQQVEER.A
3	1021.5300	1020.5227	1020.5029	0.0198	56	63	0	R.FFDTVHQQ.T
4	1043.533	n/a	n/a	n/a	n/a	n/a	n/a	n/a
5	1070.6970	1069.6897	1069.6648	0.0249	868	877	0	R.VVPAFIGVIR.A
6	1076.6600	1075.6527	1075.6502	0.0025	159	167	1	R.RYSAVLVLR.E
7	1083.6070	1082.5997	1082.6084	-0.0087	1037	1046	0	R.ILAGQEPSLR.V
8	1238.6710	1237.6637	1237.6819	-0.0182	101	110	0	R.FQQYIGTILR.G
9	1316.7110	1315.7037	1315.7149	-0.0112	610	620	0	R.HNPAYVIPHLR.K
10	1361.7040	1360.6967	1360.7139	-0.0172	996	1007	0	R.LFDSPAQPAFLR.K
11	1383.689	n/a	n/a	n/a	n/a	n/a	n/a	n/a
12	1489.7880	1488.7807	1488.8089	-0.0282	996	1008	1	R.LFDSPAQPAFLRK.S
13	1693.8220	1692.8147	1692.8624	-0.0477	584	597	0	R.TLFFALHDEQFAVR.E
14	1699.9090	1698.9017	1698.9305	-0.0287	342	356	0	K.SSITPYLDSILVHVR.E
15	2059.9400	2058.9327	2059.0156	-0.0829	39	55	1	R.ELVHIKQEMSAEQFQR.F + Oxidation (M)
16	2205.1330	2204.1257	2204.1841	-0.0584	773	791	1	K.LLGILGALDPYKYQQVEER.A
17	2254.119	n/a	n/a	n/a	n/a	n/a	n/a	n/a
18	2355.118	n/a	n/a	n/a	n/a	n/a	n/a	n/a
19	2371.118	n/a	n/a	n/a	n/a	n/a	n/a	n/a
20	2568.2570	2567.2497	2567.3456	-0.0958	562	583	2	R.QTVLTALDERFDQHLAKAENIR.T
21	2803.4380	2802.4307	2802.5142	-0.0835	436	461	0	K.VLCGEPRPLGAPYSTPIPSIPPIPK.D

Table 3. Peptide mass fingerprint for TOR^{AC}. Peaks correspond to Figure 4. **Characterisation of TOR^{AC}.** **A**, MALDI-TOF MS profile of TOR^{AC} tryptic digest. **B**, Sequence of TOR^{AC}. Sequence coverage of peptide hits returned from MASCOT (**Table**) are bold and underlined.

.A. Peaks 4, 11 and 17-19 were not found by MASCOT.

Conclusion

Here we present a procedure for obtaining the N-terminal HEAT-repeat region of the TOR protein from *C. thermophilum*. Despite advances in visualisation of the TOR complex by electron microscopy, and the solution of X-ray crystal structures of the C-terminal region of the TOR protein, the HEAT-repeat domain has so far eluded characterisation at atomic resolution. Its proposed role in binding associated proteins and recruiting partner proteins—as evidenced by biochemical approaches [6, 23], as well as recent electron microscopy studies [11, 13,

14]—illustrates the importance of this domain to the structure of the TOR complexes. Atomic resolution data will be essential for a complete understanding of the mechanistic role of this region, and its function in TOR signalling. Moreover, the capacity to purify the N-terminal region of the TOR protein will facilitate the biochemical characterization of its putative role as recruiter of TORC partners (effectors and substrates).

Acknowledgements

We wish to thank all members of the TOR Structure and Regulation Consortium for stimulating discussion and inspired collaboration. Specifically, we thank Prof. Robbie Loewith, Prof. Thanos Halazonetis, Prof. Christiane Schaffitzel, Dr. Manoel Prouteau, Dr. Taiana Oliveira, and Dr. Sotirios Sotiriou for technical advice and comments. Additionally, we are grateful to Prof. Imre Berger for Baculovirus vectors, and advice with Multibac expression. Thanks to Thomas Schalch for help with insect cell-related equipments and protocols for MALDI-TOF MS. We would like to thank Samia Barriot for advice on *C. thermophilum* culture, and Prof. Nicolas Winssinger for access to the mass spectrometer. This work was funded by a Swiss National Science Foundation Synergia Grant (Grant number CRSII3_136254), and the University of Geneva.

References

1. Heitman, J., N.R. Movva, and M.N. Hall, *Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast*. Science, 1991. **253**(5022): p. 905-9.
2. Keith, C.T. and S.L. Schreiber, *PIK-RELATED KINASES - DNA-REPAIR, RECOMBINATION, AND CELL-CYCLE CHECKPOINTS*. Science, 1995. **270**(5233): p. 50-51.
3. Loewith, R., et al., *Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control*. Mol Cell, 2002. **10**(3): p. 457-68.
4. Jacinto, E., et al., *Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive*. Nat Cell Biol, 2004. **6**(11): p. 1122-8.
5. Lempiainen, H. and T.D. Halazonetis, *Emerging common themes in regulation of PIKKs and PI3Ks*. EMBO J, 2009. **28**(20): p. 3067-73.
6. Wullschlegel, S., R. Loewith, and M.N. Hall, *TOR signaling in growth and metabolism*. Cell, 2006. **124**(3): p. 471-484.
7. Loewith, R., *A brief history of TOR*. Biochem Soc Trans, 2011. **39**(2): p. 437-42.
8. Loewith, R. and M.N. Hall, *Target of rapamycin (TOR) in nutrient signaling and growth control*. Genetics, 2011. **189**(4): p. 1177-201.
9. De Virgilio, C. and R. Loewith, *Cell growth control: little eukaryotes make big contributions*. Oncogene, 2006. **25**(48): p. 6392-415.
10. Adami, A., et al., *Structure of TOR and its complex with KOG1*. Mol Cell, 2007. **27**(3): p. 509-16.
11. Baretic, D., et al., *Tor forms a dimer through an N-terminal helical solenoid with a complex topology*. Nat Commun, 2016. **7**: p. 11016.
12. Yip, C.K., et al., *Structure of the human mTOR complex I and its implications for rapamycin inhibition*. Mol Cell, 2010. **38**(5): p. 768-74.
13. Aylett, C.H.S., et al., *STRUCTURAL BIOLOGY Architecture of human mTOR complex 1*. Science, 2016. **351**(6268): p. 48-52.
14. Gaubitz, C., et al., *Molecular Basis of the Rapamycin Insensitivity of Target Of Rapamycin Complex 2*. Molecular Cell, 2015. **58**(6): p. 977-988.
15. Yang, H., et al., *mTOR kinase structure, mechanism and regulation*. Nature, 2013. **497**(7448): p. 217-23.
16. Pan, D. and Y. Matsuura, *Structures of the pleckstrin homology domain of Saccharomyces cerevisiae Avo1 and its human orthologue Sin1, an essential subunit of TOR complex 2*. Acta Crystallogr Sect F Struct Biol Cryst Commun, 2012. **68**(Pt 4): p. 386-92.
17. Dames, S.A., et al., *The solution structure of the FATC domain of the protein kinase target of rapamycin suggests a role for redox-dependent structural and cellular stability*. Journal of Biological Chemistry, 2005. **280**(21): p. 20558-20564.
18. Choi, J., et al., *Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP*. Science, 1996. **273**(5272): p. 239-42.

19. Furuita, K., et al., *Utilization of paramagnetic relaxation enhancements for high-resolution NMR structure determination of a soluble loop-rich protein with sparse NOE distance restraints*. J Biomol NMR, 2015. **61**(1): p. 55-64.
20. Perry, J. and N. Kleckner, *The ATRs, ATMs, and TORs are giant HEAT repeat proteins*. Cell, 2003. **112**(2): p. 151-5.
21. Knutson, B.A., *Insights into the domain and repeat architecture of target of rapamycin*. J Struct Biol, 2010. **170**(2): p. 354-63.
22. Andrade, M.A. and P. Bork, *HEAT REPEATS IN THE HUNTINGTONS-DISEASE PROTEIN*. Nature Genetics, 1995. **11**(2): p. 115-116.
23. Wulschleger, S., et al., *Molecular organization of target of rapamycin complex 2*. Journal of Biological Chemistry, 2005. **280**(35): p. 30697-30704.
24. La Touche, C.J., *On a thermophile species of Chaetomium*. Transactions of the British Mycological Society, 1950. **33**(1): p. 94-IN7.
25. Millner, P.D., *Radial growth responses to temperature by 58 Chaetomium species, and some taxonomic relationships*. Mycologia, 1977. **69**(3): p. 492-502.
26. Amlacher, S., et al., *Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile*. Cell, 2011. **146**(2): p. 277-89.
27. Dimitrova, L., et al., *Structural Characterization of the Chaetomium thermophilum TREX-2 Complex and its Interaction with the mRNA Nuclear Export Factor Mex67:Mtr2*. Structure, 2015. **23**(7): p. 1246-1257.
28. Kornprobst, M., et al., *Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome*. Cell, 2016. **166**(2): p. 380-93.
29. McWilliam, H., et al., *Analysis Tool Web Services from the EMBL-EBI*. Nucleic Acids Res, 2013. **41**(Web Server issue): p. W597-600.
30. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. Mol Syst Biol, 2011. **7**: p. 539.
31. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. Anal Biochem, 1976. **72**: p. 248-54.
32. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server*. Nat Protoc, 2009. **4**(3): p. 363-71.
33. Berger, I., D.J. Fitzgerald, and T.J. Richmond, *Baculovirus expression system for heterologous multiprotein complexes*. Nat Biotechnol, 2004. **22**(12): p. 1583-7.
34. Bieniossek, C., et al., *MultiBac: expanding the research toolbox for multiprotein complexes*. Trends Biochem Sci, 2012. **37**(2): p. 49-57.
35. Bieniossek, C., et al., *Automated unrestricted multigene recombineering for multiprotein complex production*. Nat Methods, 2009. **6**(6): p. 447-50.
36. Bieniossek, C., T.J. Richmond, and I. Berger, *MultiBac: multigene baculovirus-based eukaryotic protein complex production*. Curr Protoc Protein Sci, 2008. **Chapter 5**: p. Unit 5 20.
37. Trowitzsch, S., et al., *New baculovirus expression tools for recombinant protein complex production*. J Struct Biol, 2010. **172**(1): p. 45-54.
38. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. Nature, 1970. **227**(5259): p. 680-5.
39. Rasband, W., *ImageJ*. 1997, National Institutes of Health: Bethesda, Maryland, USA.
40. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes*. Nature Protocols, 2006. **1**(6): p. 2856-2860.
41. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-3567.